

Final Report

Oxygenase-Catalyzed Biodegradation of Emerging Water
Contaminants: 1,4-Dioxane and N-Nitrosodimethylamine

SERDP Project ER-1417

February 2012

Lisa Alvarez-Cohen
University of California, Berkeley

This document has been cleared for public release



REPORT DOCUMENTATION PAGE				<i>Form Approved OMB No. 0704-0188</i>							
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.											
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.											
1. REPORT DATE (DD-MM-YYYY) 15-02-2012		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 5/2005-1/2012							
4. TITLE AND SUBTITLE Oxygenase-Catalyzed Biodegradation of Emerging Water Contaminants: 1,4-Dioxane and N-Nitrosodimethylamine				5a. CONTRACT NUMBER W912HY-05-C-0027							
				5b. GRANT NUMBER W74RDV-4349-2292							
				5c. PROGRAM ELEMENT NUMBER							
6. AUTHOR(S) Lisa Alvarez-Cohen, Ariel Grostern, Rebecca Parales, and Yinjie Tang				5d. PROJECT NUMBER ER-1417							
				5e. TASK NUMBER							
				5f. WORK UNIT NUMBER							
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Regents of the University of California Lisa Alvarez-Cohen / Civil and Environmental Engineering 726 Davis Hall, Berkeley, CA 91720-1710				8. PERFORMING ORGANIZATION REPORT NUMBER							
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Extramural Funds Accounting 2195 Hearst Ave, Room 130 MC 1103 University of California Berkeley, CA 94720-1103				10. SPONSOR/MONITOR'S ACRONYM(S) SERDP							
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) ER-1417							
12. DISTRIBUTION/AVAILABILITY STATEMENT											
13. SUPPLEMENTARY NOTES											
14. ABSTRACT 1,4-Dioxane (dioxane) and n-nitrosodimethylamine (NDMA) are emerging groundwater contaminants that are probable human carcinogens and not significantly attenuated by volatilization or sorption. The objective of this SERDP project was to identify organisms, enzymes and biochemical pathways involved in the aerobic biodegradation of these compounds. Bacterial isolates were tested for degradation of NDMA and dioxane under monooxygenase-inducing conditions and the role of monooxygenases was verified with molecular techniques by knocking out or heterologously expressing putative monooxygenase genes. The effect of the common co-contaminants 1,1,1-TCA and 1,1-DCE on dioxane degradation was tested, as was the effect of propane on NDMA degradation. Advanced analytical chemistry was used to identify the transformation intermediates of dioxane metabolism. The genome of the dioxane metabolizing actinobacterium Pseudonocardia dioxanivorans strain CB1190 was sequenced and a gene expression microarray was applied to identify genes linked to dioxane and tetrahydrofuran metabolism. Amino acid isotopomer anal											
15. SUBJECT TERMS dioxane, NDMA, monooxygenase, genome sequence, pseudonocardia, TCA, DCE, tetrahydrofuran											
16. SECURITY CLASSIFICATION OF: <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%; padding: 2px;">a. REPORT</td> <td style="width: 33%; padding: 2px;">b. ABSTRACT</td> <td style="width: 33%; padding: 2px;">c. THIS PAGE</td> </tr> <tr> <td style="text-align: center; padding: 2px;">U</td> <td style="text-align: center; padding: 2px;">U</td> <td style="text-align: center; padding: 2px;">U</td> </tr> </table>			a. REPORT	b. ABSTRACT	c. THIS PAGE	U	U	U	17. LIMITATION OF ABSTRACT UU		18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	c. THIS PAGE									
U	U	U									
			19a. NAME OF RESPONSIBLE PERSON Lisa Alvarez-Cohen		19b. TELEPHONE NUMBER (Include area code) 510-643-5969						

Reset

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

Abstract

Objective: 1,4-Dioxane (dioxane) and *n*-nitrosodimethylamine (NDMA) are emerging groundwater contaminants that are probable human carcinogens. Neither compound is significantly attenuated in the environment by volatilization or sorption processes, but a role for aerobic microbial processes in their removal has been observed. The objective of this SERDP project was to identify organisms, enzymes and biochemical pathways involved in the aerobic biodegradation of dioxane and NDMA, in order to develop a better understanding of the effects of bacterial degradation on the fate and persistence of dioxane and NDMA in the environment. This project focused specifically on oxygenase-catalyzed biodegradation of the targeted compounds.

Technical Approach: Bacterial isolates were tested for degradation of NDMA or dioxane under monooxygenase-inducing conditions. The potential role of monooxygenases in contaminant removal was confirmed by exposing cells to the monooxygenase inhibitor acetylene and monitoring for loss of activity. The role of monooxygenases was also verified with molecular techniques by knocking out or heterologously expressing putative monooxygenase genes. The effect of the common co-contaminants 1,1,1-trichloroethane and 1,1-dichloroethene on dioxane degradation was tested, as was the effect of propane on NDMA degradation. Analytical chemistry techniques were used to identify the suite of transformation intermediates of dioxane metabolism. The genome of the dioxane metabolizing actinobacterium *Pseudonocardia dioxanivorans* strain CB1190 was sequenced and a gene expression microarray was applied to identify genes linked to dioxane and tetrahydrofuran metabolism. Amino acid isotopomer analysis with ¹³C-labeled carbon substrates was used to demonstrate the activity of strain CB1190 metabolic pathways during dioxane and C2 substrate metabolism. Transformation of strain CB1190 with plasmid and transposomes was tested in order to develop a set of genetic tools for verifying the importance of particular genes in strain CB1190 metabolism.

Results: NDMA degradation in tested actinobacteria was linked to propane degradation, and in *Rhodococcus jostii* strain RHA1 directly to a propane monooxygenase. In two tested isolates, propane negatively impacted NDMA degradation. Monooxygenases were also linked to dioxane degradation in a number of tested bacterial isolates, and dioxane-transforming activity was found to be relatively common among monooxygenase-expressing bacteria. Both 1,1,1-trichloroethane and 1,1-dichloroethene inhibited dioxane degradation by bacterial isolates. The first complete pathway for the mineralization of dioxane was proposed based on identification of transformation intermediates. The complete sequence of the strain CB1190 genome was obtained and both chromosomal- and plasmid-encoded genes were induced by dioxane and THF. Strain CB1190 was shown to use the glyoxylate carboligase pathway during dioxane metabolism, which identified how energy is obtained by dioxane degradation. The plasmid-encoded monooxygenase was demonstrated to transform both dioxane and THF. Strain CB1190's large genome encodes for the utilization of a wide variety of carbon and nitrogen compounds, and annotation of some carbon utilization pathways was verified. The transformation of strain CB1190 was achieved, which will enable further development of a genetic system for this dioxane-metabolizing bacterium.

Benefits: This work highlights the importance of monooxygenases in the degradation of both NDMA and dioxane. The induction of monooxygenases in (co)metabolizing bacteria will lead to improved methods for the removal of NDMA and dioxane from contaminated groundwater.

Table of Contents

ABSTRACT	I
LIST OF TABLES	V
LIST OF FIGURES	V
LIST OF ACRONYMS AND ABBREVIATIONS	VI
KEYWORDS	VII
CONTRIBUTORS.....	VIII
1. INTRODUCTION TO SERDP ER1417	1
DIOXANE AND NDMA – STATEMENT OF NEED AND OBJECTIVES.....	1
BACKGROUND.....	2
Environmental Relevance of dioxane and NDMA	2
Existing knowledge of dioxane and NDMA biotransformation prior to start of project.....	3
Formation of the hypothesis: microbial monooxygenases and mammalian P-450 enzymes	3
2. ENZYMES AND KINETICS OF BACTERIAL NDMA DEGRADATION	5
INTRODUCTION.....	5
RESULTS AND DISCUSSION	6
CONCLUSION.....	9
3. BACTERIAL STRAIN EVALUATION, DEGRADATION KINETICS, PATHWAY DETERMINATION AND CO-CONTAMINANT INHIBITORY EFFECTS FOR DIOXANE DEGRADATION.....	10
INTRODUCTION.....	10
RESULTS AND DISCUSSION	11
CONCLUSION.....	15
4. SEQUENCING AND ANALYSIS OF THE GENOME OF THE DIOXANE DEGRADING ACTINOMYCETE <i>PSEUDONOCARDIA DIOXANIVORANS</i> STRAIN CB1190.....	17
INTRODUCTION.....	17
MATERIALS AND METHODS.....	18
Culture conditions	18
Genomic DNA extraction	18
Genomic sequencing and assembly	19
Genome annotation and analysis.....	19
RESULTS	20
Genome properties and features of <i>P. dioxanivorans</i> strain CB1190.....	20
Comparative genomics of <i>Pseudonocardiaceae</i> family	20
Mobile genetic elements	23
Signal transduction systems	23
Transport systems.....	25
Secretion systems	26
Chaperone proteins	26
Nitrogen metabolism	26
General carbon metabolism.....	27
Carbon-fixation pathways.....	27
Monooxygenases.....	28
Dioxygenases.....	30
DISCUSSION.....	31
CONCLUSION.....	32
5. TRANSCRIPTIONAL ANALYSIS OF <i>PSEUDONOCARDIA DIOXANIVORANS</i> STRAIN CB1190 DIOXANE DEGRADATION.....	33

INTRODUCTION.....	33
MATERIALS AND METHODS.....	34
Culture growth	34
Cell harvesting and RNA extraction for transcription studies.....	34
Microarray design and analyses.....	35
Quantitative RT-PCR	36
Cloning and expression of strain CB1190 genes.....	38
Glyoxylate carboligase assay.....	38
Growth with 1,4-[U- ¹³ C]dioxane	38
Analytical methods	38
RESULTS AND DISCUSSION	39
<i>In silico</i> analysis of dioxane degradation pathway genes.....	39
Gene expression during growth on dioxane, glycolate and pyruvate.....	40
Upregulation of a multi-component monooxygenase gene cluster during dioxane metabolism.....	42
Genes potentially contributing to transformation of C ₄ dioxane metabolites	44
Genes potentially contributing to transformation of C ₂ dioxane metabolites to glyoxylate	44
Glyoxylate metabolism during dioxane degradation by strain CB1190.....	45
Amino acid isotopomer analysis to identify routes for dioxane carbon assimilation.....	47
Energy generation in strain CB1190 dioxane metabolism.....	49
6. TRANSCRIPTOMICS OF THE DEGRADATION AND INVESTIGATION OF MONOOXYGENASE ACTIVITIES IN <i>PSEUDONOCARDIA DIOXANIVORANS</i> STRAIN CB1190.....	50
INTRODUCTION.....	50
METHODS	51
Chemicals	51
Culture conditions	51
Cell harvesting and RNA isolation for transcriptional studies.....	51
Analytical methods	51
Transcriptomics microarray analysis	52
Quantitative RT-PCR analyses.....	52
<i>thm</i> gene cluster transcriptional analysis.....	53
Cloning and expression of strain CB1190 THF monooxygenase genes in <i>R. jostii</i> strain RHA1	53
HEAA transformation inhibition assay	54
RESULTS AND DISCUSSION	54
Transcriptomics of <i>P. dioxanivorans</i> strain CB1190 growth on THF and succinate.....	54
Induction of <i>thmA</i> gene expression by dioxane and THF	58
Functional activity of <i>thmADBC</i> -expressing <i>R. jostii</i> RHA1 clones.....	60
Growth of <i>P. dioxanivorans</i> strain CB1190 on HEAA	62
Testing the effect of acetylene exposure on HEAA degradation.....	63
CONCLUSION.....	64
7. USING METABOLOMICS ANALYSES TO ASSIST GENOME ANNOTATION OF <i>PSEUDONOCARDIA DIOXANIVORANS</i> STRAIN CB1190.....	65
INTRODUCTION.....	65
Strain CB1190 and genome annotation.....	65
Ethylmalonyl-CoA pathway and glyoxylate cycle pathway	65
Glycine cleavage system and glycine-glyoxylate pathway	66
Research questions, hypotheses and important results.....	66
MATERIALS AND METHODS.....	66
Chemicals	66
Bacterial strain and culture conditions	66
Analytical methods	67
RNA extraction for transcription studies	67
Reverse transcript-PCR.....	67
Cell-free extracts and enzyme measurement	67
RESULTS AND DISCUSSION	67

¹³ C labeled pyruvate to evaluate annotation of central pathways	67
Demonstration of acetate assimilation using stable isotopic analyses.....	68
Demonstration of glycine assimilation in strain CB1190 using stable isotopic analyses.....	70
CONCLUSION.....	72
8. TOWARDS DEVELOPMENT OF A GENETIC SYSTEM FOR <i>PSEUDONOCARDIA DIOXANIVORANS</i>	
STRAIN CB1190.....	74
INTRODUCTION.....	74
MATERIALS AND METHODS.....	74
Antibiotic resistance testing of CB1190.....	74
Conjugative matings with <i>E. coli</i> to introduce plasmids into CB1190	74
Transposon mutagenesis of CB1190	74
Preparation of electrocompetent cells	75
Electroporation of CB1190.....	75
RESULTS AND DISCUSSION	75
CONCLUSION.....	76
9. IMPLICATIONS OF RESEARCH FOR NDMA AND DIOXANE BIOREMEDIATION	77
10. LITERATURE CITED	78
APPENDIX 1. LIST OF SCIENTIFIC PUBLICATION	89
ARTICLES IN PEER-REVIEWED JOURNALS	89
CONFERENCE ABSTRACTS.....	89
PH.D. DISSERTATIONS	90

List of Tables

Table 3.1. Bacterial strains tested for the ability to degrade dioxane.	11
Table 4.1. Genome feature of <i>P. dioxanivorans</i> strain CB1190.	20
Table 4.2. Features of sequenced <i>Pseudonocardia</i> genomes ^a	22
Table 4.3. Reciprocal best blast hits between <i>Pseudonocardiaceae</i>	24
Table 5.1. SYBR Green and TaqMan primers used in qRT-PCR analyses. ^{a,b}	37
Table 5.2. Strain CB1190 genes on microarray up-regulated during growth with dioxane ^a	41
Table 5.3. Normalized relative expression levels for selected strain CB1190 genes analyzed by qRT-PCR.	43
Table 6.1 Genes up-regulated on THF but not succinate, relative to pyruvate.	55
Table 7.1. Amino acids ¹³ C stable isotopic labeling profiles used to identify acetate assimilation pathways in strain CB1190.	68
Table 7.2A. Amino acids ¹³ C stable isotopic labeling profiles used to identify glycine assimilation pathways in strain CB1190.	71
Table 7.2B. Amino acids ¹³ C stable isotopic labeling profiles used to identify glycine assimilation pathways in strain CB1190.	72
Table 8.1. Results of CB1190 electroporation with pMyVec1.	76

List of Figures

Figure 2.1. Constitutive removal of NDMA occurs at a fraction of the propane-induced rate.	6
Figure 2.2. Effect of propane on transcription of three aliphatic monooxygenase components as quantified by RT-qPCR (white bars) and spotted microarray (grey bars).	7
Figure 2.3. Inhibitory effect of propane on NDMA degradation.	8
Figure 3.1. Monod plot of dioxane degradation by (a) <i>Pseudonocardia dioxanivorans</i> CB1190 and (b) <i>Pseudonocardia benzenivorans</i> B5 over a range of concentrations.	12
Figure 3.2. Final distribution of ¹⁴ C derived from the degradation of dioxane after 48 h incubation.	13
Figure 3.3. Observed biodegradation pathway of dioxane by monooxygenase-expressing bacteria in this study.	14
Figure 3.4. Degradation of dioxane by CB1190 in the presence of various concentrations of TCA.	15
Figure 4.1. Bacterial multicomponent monooxygenases (BMMs) in strain CB1190.	29
Figure 4.2. THF MO gene clusters from strain CB1190 and <i>P. tetrahydrofuranoxydans</i>	30
Figure 5.1. Proposed pathways and enzymes involved in dioxane metabolism in strain CB1190.	35
Figure 5.2. Strain CB1190 chromosomal regions of interest implicated in glycolate and glyoxylate transformations during dioxane metabolism.	46
Figure 5.3. Glyoxylate carboligase activity in strain CB1190 and strain RHA1/pTip cell-free extracts.	47
Figure 5.4. Proposed assimilation of unlabeled carbon during growth of strain CB1190 with uniformly ¹³ C-labeled dioxane.	48
Figure 6.1. RT-PCR analysis of THF monooxygenase gene expression.	59
Figure 6.2. Induction of <i>thmA</i> gene expression in glucose-grown strain CB1190.	60

Figure 6.3. Transformation of dioxane to HEAA by heterologous strain CB1190 monooxygenase.	61
Figure 6.4. Removal of THF by heterologous strain CB1190 monooxygenase.	62
Figure 6.5. Degradation of HEAA by strain CB1190.	63
Figure 6.6. Test of acetylene on HEAA removal by strain CB1190.	64
Figure 7.1. Schematic of the ethylmalonyl-CoA pathway in CB1190 with possible genes encoding the enzyme assigned to each step.	69
Figure 7.2. ¹³ C stable isotopic analysis method used to differentiate glycine assimilation pathways.	71
Figure 7.3. Schematic of the glycine cleavage system pathway in strain CB1190 with possible genes encoding the enzyme assigned to each step.	72

List of Acronyms and Abbreviations

3-HP	3-hydroxypropionate
ABC	ATP-binding cassette
AlkMO	alkane monooxygenase
AMS	ammonium mineral salts
ANOVA	analysis of variance
ATP	adenosine triphosphate
BMM	bacterial multi-component monooxygenase
CBB	Calvin-Benson-Bassham
CDS	coding sequence
DCE	1,1-dichloroethene
Dioxane	1,4-dioxane
ECF	extracytoplasmic function
EPA	Environmental Protection Agency
FAD	flavin adenine dinucleotide
FID	flame ionization detector
GC-MS	gas chromatograph-mass spectrometer
GCS	glycine cleavage system
HEAA	2-hydroxyethoxyacetic acid
HILIC	hydrophilic interaction chromatograph
HK	histidine kinase
HPA	4-hydroxyphenylacetate
HPLC	high pressure liquid chromatography
JGI	Joint Genome Institute
K_s	Monod half-velocity constant
LB	Luria broth
Mb	megabase
MGE	mobile genetic element
MMO	methane monooxygenase
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NDMA	<i>n</i> -nitrosodimethylamine
ORF	open reading frame
PEP	phosphoenolpyruvate
Pkinase	protein kinase

PP	pentose phosphate
PDD	pulsed discharge detection
PrMO	propane monooxygenase
qRT-PCR	quantitative reverse transcription-PCR
RR	response regulator
RT-PCR	reverse transcription-PCR
SDS	sodium dodecyl sulfate
sMMO	soluble methane monooxygenase
SON	Statement of Need
strain CB1190	<i>Pseudonocardia dioxanivorans</i> strain CB1190
strain RHA1	<i>Rhodococcus jostii</i> strain RHA1
TCA	1,1,1-trichloroethane
TCA	tricarboxylic acid
THF	tetrahydrofuran
THFMO	THF monooxygenase
TMO	toluene monooxygenase
T4MO	toluene-4-monooxygenase
V_{\max}	Monod maximum transformation rate

Keywords

NDMA, n-nitrosodimethylamine, 1,4-dioxane, *Rhodococcus*, *Pseudonocardia*, strain CB1190, biodegradation, bioremediation, monooxygenase, 1,1,1-trichloroethane, 1,1,1-TCA, 1,1-dichloroethene, 1,1-DCE, propane, THF, tetrahydrofuran, enzyme kinetics, qRT-PCR, microarray, glyoxylate, glycine, acetate, genome

Contributors

This report was prepared by:

Ariel Grostern, Department of Environmental & Civil Engineering, UC Berkeley

Contributors to work in this report:

University of California, Berkeley – Department of Civil & Environmental Engineering

Lisa Alvarez-Cohen

Ariel Grostern

Shaily Mahendra

Christopher M. Sales

Jonathan O. Sharp

Wei-Qin Zhuang

University of California, Berkeley – Department of Plant & Microbial Biology

Onur Erbilgin

University of California, Berkeley – Department of Chemical Engineering

Christopher J. Petzold

Edward E. Baidoo

Jay D. Keasling

University of California, Davis – Department of Microbiology

Rebecca Parales

Juan Parales

Washington University in St. Louis – Department of Energy, Environmental & Chemical Engineering

Yinjie Tang

Xueyang Feng

University of British Columbia – Department of Microbiology & Immunology

Justin C. LeBlanc

Jie Liu

Lindsay D. Eltis

William W. Mohn

Joint Genome Institute

Tanja Woyke

Matt Nolan

Alla Lapidus

Alexander Sczyrba

Galina Ovchinnikova

Los Alamos National Laboratory

Lynne A. Goodwin

Olga Chertkov

1. Introduction to SERDP ER1417

Dioxane and NDMA – Statement of Need and Objectives

This research examined the biodegradation of the emerging water contaminants 1,4-dioxane (dioxane) and N-nitrosodimethylamine (NDMA). Both dioxane and NDMA are probable human carcinogens and confirmed animal carcinogens. Neither is significantly attenuated in the environment by volatilization or sorption processes. Due to its widespread use as a solvent stabilizer, dioxane is frequently found commingled with chlorinated solvents at U.S. Department of Defense and Department of Energy sites. NDMA is found as a degradation byproduct in proximity to aerospace facilities that used hydrazine-based rocket fuel.

Although the carcinogenic threats of dioxane and NDMA have been understood for many years, they have not historically been considered important water quality issues, mostly due to lack of awareness about their potential occurrence in drinking water supplies. However, with recent advances in analytical methods and growing public awareness of their occurrence in drinking water supplies, dioxane and NDMA are emerging as important water contaminants. Consequently, a better understanding of the effects of bacterial degradation on the fate and persistence of dioxane and NDMA in the environment is needed.

This research responded to SON number CUSON-05-01, with the overall goal of identifying organisms, enzymes and biochemical pathways involved in the aerobic biodegradation of dioxane and NDMA. The work was then extended to examine the molecular basis for dioxane degradation by the actinomycete *Pseudonocardia dioxanivorans* strain CB1190. This work aimed to provide a mechanistic understanding of the degradation of dioxane and NDMA by aerobic microorganisms and as such to provide a foundation for the bioremediation of these contaminants in natural and engineered systems.

The specific objectives of the original piece of this project were:

1. Test oxygenase-expressing bacteria for NDMA and dioxane degradation
2. Evaluate the mechanisms and pathways of NDMA and dioxane degradation
3. Quantify the kinetics of NDMA and dioxane oxidation reactions at both moderate and trace concentrations
4. Investigate the impacts of common co-contaminants and inducing substrates on NDMA and dioxane degradation kinetics

Specific objectives of the extension were:

5. Sequence and analyze the genome of the dioxane metabolic degrader *P. dioxanivorans* strain CB1190.
6. Develop a whole genome expression microarray for strain CB1190 and apply to identify genes important in metabolism and response to environmental stresses
7. Analyze metabolomics and fluxomics for strain CB1190 metabolism
8. Develop a genetic system for strain CB1190

These objectives were used to guide experimental activities, which are described and summarized in sections 2 through 8 of this document, while sections 9 and 10 describe the implications of the research and the cited literature, respectively.

Background

Environmental Relevance of dioxane and NDMA

Dioxane is classified as a probable human carcinogen and a confirmed animal carcinogen (56), and exposure to dioxane causes damage to the liver, kidneys and nervous system (30). Dioxane is a cyclic organic ether that is widely used as a stabilizer for chlorinated solvents such as 1,1,1-trichloroethane (TCA). Dioxane is also used as a solvent in paper and textile processing, and in the manufacture of several organic chemicals (181). As a result of its widespread use and improper disposal, dioxane has been detected in surface waters and groundwaters across the United States, usually as a co-contaminant of TCA. In an aquifer in the San Gabriel Basin, an eight mile plume was found that was believed to have several dozen sources (98). Dioxane has also been identified as a frequent contaminant in the lower Mississippi River (59). In Washtenaw County, Michigan, groundwater aquifers at a medical filters manufacturing site are heavily contaminated with dioxane at concentrations up to 30,000 µg/L (166). At present, there is no federal maximum contaminant level for dioxane. Several states have adopted water quality standards and/or guidelines for dioxane levels ranging from 3 to 85 µg/L (32, 98). Sorption onto activated carbon and air-stripping techniques have proven to be ineffective in removing dioxane from water because of its high solubility and low vapor pressure (Henry's Law constant is $5 \times 10^{-6} \text{ atm}\cdot\text{m}^3\text{mol}^{-1}$ at 20° C). Dioxane does not sorb readily to soils ($\log K_{ow} = -0.27$). Consequently, dioxane groundwater plumes generally expand faster than associated chlorinated solvent plumes in aquifers (58). Photocatalytic oxidation is possible, but is often expensive and requires ex-situ remediation with pump-and treat operations (54, 144). Aitchison et al. (1) have reported phytoremediation of dioxane by hybrid poplar trees. In their laboratory studies, there was rapid uptake of dioxane from aqueous solution as well as from soil, and most of the dioxane was transpired from leaf surfaces instead of being transformed.

NDMA is also classified as a probable human and confirmed animal carcinogen (7). The US EPA has established a cleanup level of 0.7 ng/L for NDMA in groundwater (160), based on a risk assessment target of an increased lifetime cancer risk of 10^{-6} in drinking water. The environmental occurrence of this compound is regulated to a concentration about one-one thousandth that of dioxane with a California Department of Health action level of 10 ng/L (22). NDMA is highly water-soluble, and though volatilization can play a role in surface waters, given its low Henry's constant ($2.6 \times 10^{-7} \text{ atm}\cdot\text{m}^3\text{mol}^{-1}$ at 20° C), volatilization should be an insignificant process in groundwater (97, 107). NDMA does not significantly sorb to activated carbon, and it is highly mobile in soil regardless of the organic carbon, humic acid, pH, clay, or cation exchange capacity of the soil (29, 44, 63). NDMA's presence in the environment has been linked to aerospace facilities through the decomposition of hydrazine-based rocket fuels (21) and to the discharge of chlorine-based disinfected water and wastewater (21, 100, 109). Its persistence in groundwater aquifers has been responsible for the closure of at least two municipal drinking water wells in Orange County, California (109). NDMA groundwater concentrations as high as 400,000 ng/L on-site and 20,000 ng/L off-site have been detected at a rocket engine testing facility in Sacramento County, California, that used unsymmetrical dimethylhydrazine (UDMH)-based rocket fuel. These extreme concentrations prompted closure of downgradient drinking water wells at this site (21). NDMA has also been detected in groundwater downgradient of a rocket engine testing facility in the San Gabriel Valley, California, at concentrations up to 3,000 ng/L (21) and groundwater containing 350 ng/L NDMA has been documented at the Rocky Mountain Arsenal in Commerce City, Colorado.

Existing knowledge of dioxane and NDMA biotransformation prior to start of project

Bioremediation may be a cost-effective alternative for the treatment of dioxane and NDMA; however, prior to the start of this project, little was known about the microbial biodegradation of these compounds.

There was some evidence for biodegradation of dioxane by aerobic microorganisms. Bacterial strains, *Mycobacterium vaccae* and *Rhodococcus* sp., as well as a fungus *Aureobasidium pullmans*, were shown to degrade dioxane cometabolically, but dioxane biodegradation did not support growth of these organisms (13, 19, 115). Aerobic biodegradation of high concentrations of dioxane by mixed cultures in industrial sludge has also been reported (125, 127). Parales et al. (112) showed that a nocardioform culture, CB1190, was capable of growing on dioxane. One study showed dioxane degradation by cometabolism using tetrahydrofuran as the primary substrate; however, tetrahydrofuran is a toxic chemical itself and is thus unsuitable for enhancing in-situ bioremediation (180).

The literature associated with the microbial degradation of NDMA was even more sparse. In the non-microbial world, the microsomal fractions of tulip bulbs (145), spinach, and lettuce (29) had all been shown to metabolize NDMA. In addition, many higher organisms including mammals possess enzymes that have the ability to degrade NDMA (74, 157). In the microbial world, undefined consortia grown aerobically had demonstrated the ability to degrade NDMA in laboratory incubations (44, 63, 151). Biological activity was implicated for the disappearance of NDMA at a field site in Colorado (44). Kaplan and Kaplan (63) observed that an activated carbon trickling filter acted as a growth matrix for NDMA degrading organisms while a sterilized column did not remove the NDMA. However, attempts to isolate bacterial strains from these NDMA-active columns failed. In fact, there had been no reports in the literature of pure cultures capable of growth on NDMA or even cometabolic degradation of this compound.

Formation of the hypothesis: microbial monooxygenases and mammalian P-450 enzymes

As described above, previous studies provided evidence that dioxane can be biodegraded aerobically by both metabolic and cometabolic reactions; however, these studies examined only a small pool of potential microorganisms and the responsible enzymes and pathways had not been elucidated. Similarly, biological activity had been implicated in the disappearance of NDMA in aerobic microcosms, but adequate attempts had not been made to identify the responsible organisms, determine degradation pathways, characterize the responsible enzymes, or quantify the reaction kinetics. Indeed, attempts to isolate bacterial strains capable of growth on NDMA had failed.

Though the literature involving microbial metabolism of dioxane and NDMA was sparse, the metabolism in mammalian systems of both compounds had been extensively studied due to concerns about their carcinogenicity (75, 157, 170, 177). Evidence from animal studies indicated that biotransformation of these compounds likely involves NADH-dependent oxidation by cytochrome P-450 monooxygenase enzymes. Monooxygenases are a generic class of enzymes that catalyze the addition of one atom of molecular oxygen directly into an organic substrate. The second oxygen atom is reduced to water either by the substrate itself or by a cosubstrate reductant such as NADH. All monooxygenases contain a cofactor, a transition metal, flavin or pteridine, which overcomes the spin-forbidden reaction between the paramagnetic dioxygen and the singlet carbon in organic compounds. A monooxygenation is separated into two steps. First, two reducing equivalents generated by the oxidation of

NAD(P)H are transferred to the cofactor which is reduced. In the second step, dioxygen couples with the reduced cofactor to form activated oxygen radicals which subsequently hydroxylate the substrate (50).

A variety of cytochrome P-450s have been identified from prokaryotes, yeast, fungi, plants, insects, and mammals (103). In prokaryotes, monooxygenases play significant roles in catabolic pathways. Mammalian monooxygenases are primarily involved in the hydroxylation of steroids, in the biosynthesis of vitamin D and neurotransmitters, and in the detoxification of poisonous compounds (50). Eukaryotic cytochrome P-450's located in the endoplasmic reticulum and mitochondria are membrane-bound, while most bacterial P-450's are soluble. An evolutionary tree representing the divergent evolution of prokaryotic and eukaryotic P-450 monooxygenases based upon a comparison of the primary structures of 154 members of the cytochrome P-450 superfamily has been described (103) and mechanisms such as genetic transfer mediated by plasmids, chromosomal gene mobilization, transduction, transformation, and transposition have been proposed to explain the evolution of cytochrome P-450 monooxygenases among prokaryotic and eukaryotic organisms. Thus monooxygenase enzymes, though found in very different types of organisms, evolved from a common ancestor and for this reason share common traits.

It is our hypothesis that since bacterial monooxygenase enzymes are similar in structure, function and reaction mechanisms to mammalian P-450 enzymes, a class of these bacterial monooxygenases will have the ability to degrade dioxane and NDMA analogously to their mammalian counterparts.

2. Enzymes and kinetics of bacterial NDMA degradation

This is a summary of work published in:

1. Sharp, J. O., C. M. Sales, J. C. LeBlanc, J. Liu, T. K. Wood, L. D. Eltis, W. W. Mohn, and L. Alvarez-Cohen. 2007. An inducible propane monooxygenase is responsible for N-Nitrosodimethylamine degradation by *Rhodococcus* sp. strain RHA1. *Appl. Environ. Microbiol.* **73**:6930-6938.
2. Sharp, J. O., C. M. Sales, and L. Alvarez-Cohen. 2010. Functional characterization of propane-enhanced N-nitrosodimethylamine degradation by two actinomycetales. *Biotechnol. Bioeng.* **107**:924-932.

Introduction

Microorganisms grown on substrates such as propane, methane, and toluene have been shown to rapidly oxidize NDMA in the laboratory (39, 138). In these cases, evidence from inhibition and induction experiments along with observations of requisite oxygen consumption suggests that propane monooxygenases (PrMO), soluble methane monooxygenases (sMMO), and toluene monooxygenases (TMO) are most likely involved in these transformations. In addition, experiments with *Escherichia coli* clones expressing TMO inserts confirmed the role of toluene 4-monooxygenase (T4MO) in NDMA oxidation, while cupric selection for soluble rather than particulate MMO confirmed the role of sMMO (138). The involvement of PrMO is less understood, as the traditional boundary can blur between enzymes oxidizing gaseous and liquid n-alkanes. Liquid alkanes are typically oxidized by alkane monooxygenases (AlkMO), but AlkMO can be induced by propane in some bacteria but not in others (47, 86). Regardless of the class of monooxygenase involved, NDMA is transformed with little observed benefit to the cells and no evidence of cellular growth, despite the production of oxidized products, including formaldehyde, methylamine, and methanol, that can be incorporated into primary metabolic pathways (39, 136, 176). Limited evidence for metabolism suggests that non-energy-generating transformations, such as cometabolic oxidation reactions, play an important role in the biological attenuation of NDMA.

Rhodococci are soil heterotrophs of the order *Actinomycetales* with a noted diversity of functional enzymatic activities (45, 80). Collectively, this order is biologically and economically significant for the production of a diverse array of enzymes involved in the production of commercial secondary metabolites, antibiotics, and the metabolism of xenobiotic compounds for environmental and industrial applications (104). Global genomic, transcriptomic, and functional analyses of *Rhodococcus jostii* strain RHA1 reveal tremendous enzymatic diversity with the potential to grow on a wide variety of aromatic compounds, carbohydrates, nitriles, and steroids as the sole carbon and energy sources (see reference (96) and references therein). Indeed, the genome of strain RHA1 is predicted to encode over 200 oxygenases, including both PrMO and AlkMO gene clusters. The genetic blueprint provided by the annotated genome and the development of a corresponding global microarray facilitate the identification of genes responsible for physiological traits of RHA1, especially for growth and enzymatic activity. For this reason, strain RHA1 was selected as a model organism to better understand the genetics and biochemistry of NDMA transformation.

The goal of this work was to determine the role of PrMO activity in the transformation of NDMA by strain RHA1. Similar analyses were then extended to other actinobacteria, *Rhodococcus* sp. strain RR1 and *Mycobacterium vaccae* strain JOB-5.

Results and Discussion

We first tested the ability of strain RHA1 to degrade NDMA. Cells grown in rich medium (LB or soy broth) or in pyruvate-amended minimal medium degraded NDMA at a rate of $0.04 \pm 0.01 \mu\text{g NDMA mg protein}^{-1}$. In contrast, propane-grown RHA1 cells removed NDMA at rates approximately 500-fold higher. The kinetics of NDMA removal are represented in Fig. 2.1. Monod kinetic parameters for NDMA removal by propane-grown RHA1 were determined: the maximum NDMA removal rate (V_{max}) was $18 \pm 3 \mu\text{g NDMA mg protein}^{-1} \text{ h}^{-1}$ and the half-saturation constant (K_S) was $20 \pm 17 \mu\text{g NDMA L}^{-1}$.

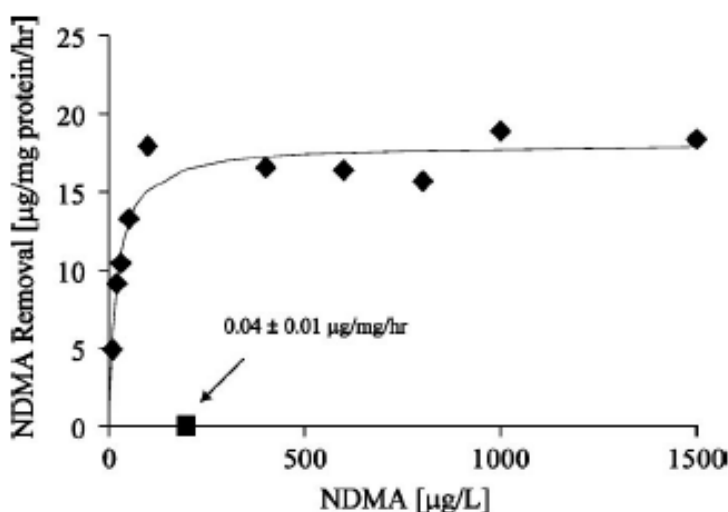


Figure 2.1. Constitutive removal of NDMA occurs at a fraction of the propane-induced rate. A Monod kinetic model (line) is fit to NDMA removal rates measured after RHA1 was grown on propane (diamonds). The constitutive NDMA degradation rate at 200 $\mu\text{g/L}$ (square) represents average removal after independent growth on three noninducing substrates (pyruvate, LB medium, or soy broth).

In order to understand the role of propane in the induction of NDMA removal by strain RHA1, we investigated the effect of propane on gene expression using a custom strain RHA1 spotted microarray. Relative to growth on pyruvate, 45 genes were up-regulated in response to propane, with the greatest fold increase (125) observed for *prmA* (ro00441), encoding the large hydroxylase subunit of PrMO. The *prmA* gene is part of a 13-gene cluster of which seven additional genes had expression ratios greater than 10. In contrast, a gene cluster encoding alkane monooxygenase (AlkMO), which has been implicated in propane transformation in some bacteria, was not significantly differentially expressed in propane-grown RHA1. These microarray results were then validated with RT-qPCR (Fig. 2.2).

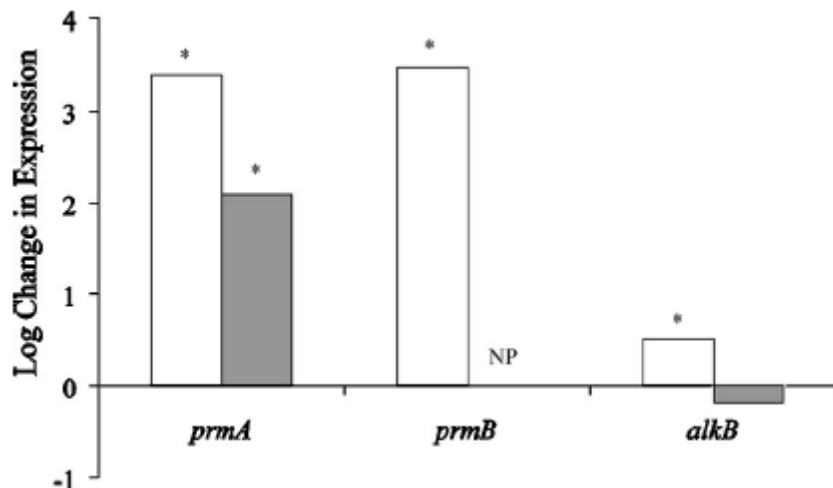


Figure 2.2. Effect of propane on transcription of three aliphatic monooxygenase components as quantified by RT-qPCR (white bars) and spotted microarray (grey bars). The microarray did not code for *prmB* (NP), preventing its quantification by that method. Asterisks denote that the propane-grown values are statistically different from those of the pyruvate-grown controls based on Student's t test.

To confirm that PrMO is the primary catalyst for NDMA in oxidation in RHA1, knockout mutant strains were generated with deletions in *prmA* and *alkB* (encoding a subunit of AlkMO). Knockout of each gene target did not affect growth of RHA1 on rich LB medium. However, while wild-type RHA1 and the *alkB* mutant grew on propane-amended minimal medium, the *prmA* mutant did not grow on propane. When wild-type or the mutant strain of RHA1 was grown on LB medium and washed cells were exposed to NDMA, the wild-type and *alkB*-mutant strains removed NDMA to below detection limits in 4 h, whereas the *prmA* mutant did not significantly remove any NDMA over 19 h.

This work demonstrated that the PrMO gene cluster in *Rhodococcus jostii* strain RHA1 encoded both for growth on propane and removal of NDMA. Due to low-level NDMA removal when RHA1 is grown with LB or soy broth rich medium, it is apparent that PrMO is constitutively expressed and is upregulated when RHA1 is exposed to propane.

We next extended our work examining the correlation of propane and NDMA degradation to two other actinobacteria, *Rhodococcus* sp. RR1 and *Mycobacterium vaccae* strain JOB-5. Each strain was first grown on propane and then the kinetics of NDMA removal were determined by exposing washed cells to varying concentrations of NDMA. The maximum NDMA removal rates V_{\max} were relatively similar (44 and 28 $\mu\text{g NDMA mg protein}^{-1} \text{ h}^{-1}$ for strains RR1 and JOB-5, respectively), whereas the half-saturation constants K_S differed by two orders of magnitude (36 and 2,200 $\mu\text{g NDMA L}^{-1}$ for strains RR1 and JOB-5, respectively). When grown on soy broth rich medium, strain RR1 had low constitutive NDMA removal activity (like strain RHA1 shown above), whereas strain JOB-5 lacked such constitutive activity. Cells of both RR1 and JOB-5 pre-grown with propane were able to remove low concentrations of NDMA, degrading 100 ng NDMA L^{-1} to below the limit of detection (20 ng/L).

The effects of propane on NDMA degradation rates were then determined. For both strains RR1 and JOB-5, propane inhibited NDMA in proportion to concentration, but inhibition kinetics analysis revealed complex inhibition kinetics that do not correspond with classical enzymatic models. By using an alternative approach (142), we were able to compare the inhibitory effects of propane on NDMA removal by the two strains, and we found that JOB-5

was significantly more sensitive than RR1 to the presence of propane (Fig. 2.3). For example, when amended with 200 $\mu\text{g NDMA L}^{-1}$, RR1 required 7700 $\mu\text{g propane L}^{-1}$ to decrease the NDMA removal rate by 50%, whereas JOB-5 required only 120 $\mu\text{g propane L}^{-1}$ to achieve the same effect.

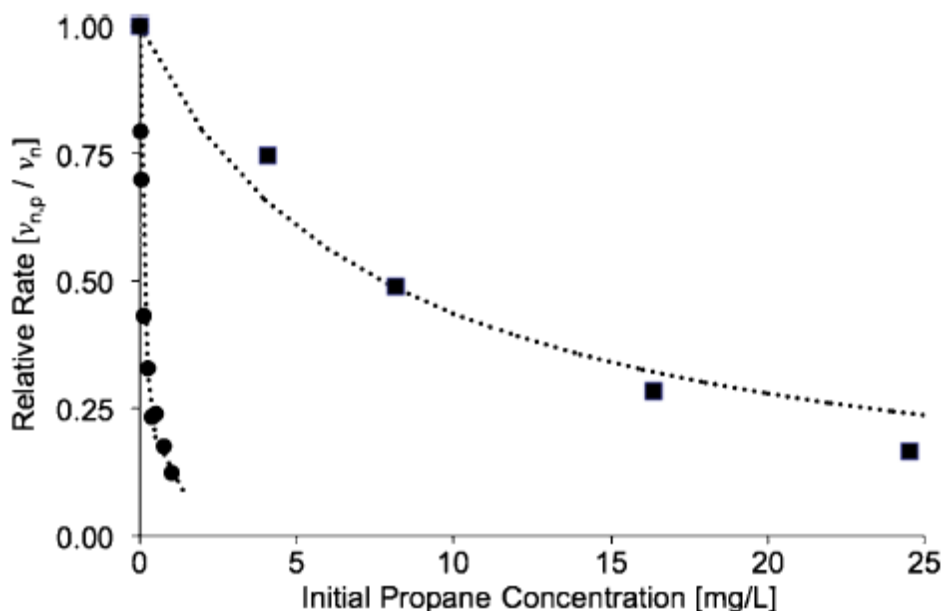


Figure 2.3. Inhibitory effect of propane on NDMA degradation. Initial relative rates for the degradation of 200 mg/L NDMA was normalized as degradation rate ($v_{n,p}$) in the presence of a given propane concentration divided by the rate (v_n) in the absence of propane. The concentration of propane in the dissolved solution was adjusted by the Henry's constant to account for liquid phase partitioning. Squares = relative rate of NDMA degradation for strain *Rhodococcus* sp. RR1; circles = relative rate of NDMA degradation for strain *Mycobacterium vaccae* JOB-5.

To gain insight into the presence of propane and alkane monooxygenases in strain RR1 and JOB-5, we first developed a *prmA* degenerate primer set based on annotated *prmA* sequences in the NCBI Entrez Protein database and attempted to amplify putative *prmA* from DNA extracted from these two strains. A PCR product of the expected size (~1300 bp) was obtained only for strain RR1, and the *prmA* fragment was found to be closely related to the strain RHA1 *prmA* sequence (91% nucleotide identity and 96% amino acid identity). We then applied *alkB* degenerate primers to DNA from each strain and obtained an expected 343 bp fragment from both strain. Sequence analyses revealed that RR1 possessed one *alkB* homologue while JOB-5 contained at least two distinct copies, which share 60% nucleotide identity.

We examined the effect of the presence of propane on the expression of genes identified above for strains RR1 and JOB-5. In RR1, *prmA* was induced 70-fold relative to controls not exposed to propane, whereas *alkB* transcription levels did not change significantly. Similarly, the expression of both *alkB* homologue in JOB-5 was not significantly changed due to the presence of propane (Fig. 2.4).

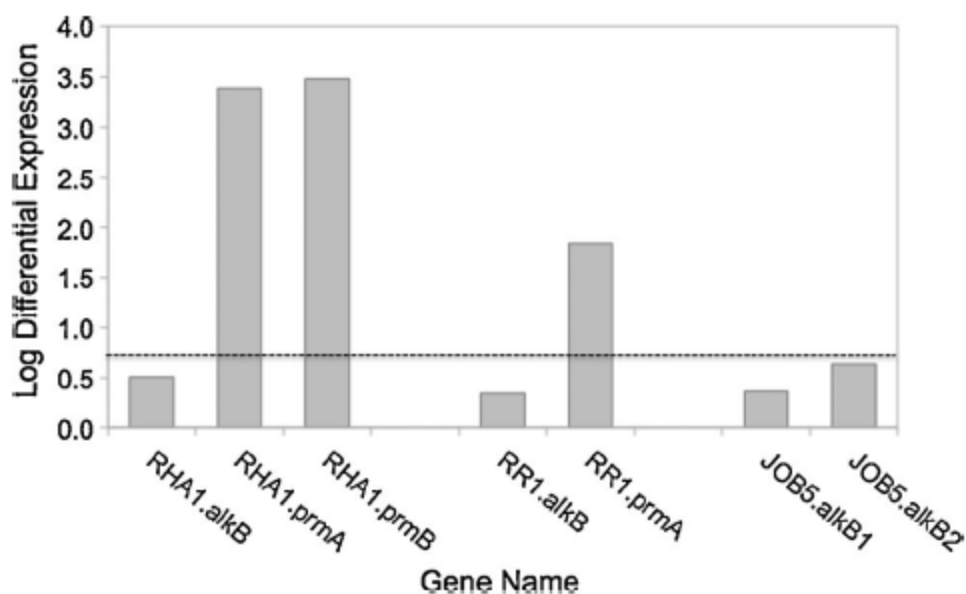


Figure 2.4. Difference in expression for candidate propane oxidizing genes when contrasting propane and pyruvate grown cultures. The dashed line represents a fivefold differential expression threshold. Expression data for strain RHA1 is from Sharp et al. (137).

Conclusion

This work established a role for PrMO in NDMA degradation by some actinobacteria. The *Rhodococcus* and *Mycobacterium* tested in this study are common soil bacteria of the order *Actinomycetales* that have previously been shown to metabolize propane. The orders of magnitude differences in transformation rates between constitutive and promoted conditions reported here and in prior studies (39, 137) suggest that the addition of propane as an inducing substrate could substantially enhance the bioremediation of NDMA in contaminated soils.

3. Bacterial strain evaluation, degradation kinetics, pathway determination and co-contaminant inhibitory effects for dioxane degradation

This is a summary of work published in:

1. Mahendra, S., and L. Alvarez-Cohen. 2006. Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. *Environ. Sci. Technol.* **40**:5435-5442.
2. Mahendra, S., C. J. Petzold, E. E. Baidoo, J. D. Keasling, and L. Alvarez-Cohen. 2007. Identification of the intermediates of *in vivo* oxidation of 1,4-dioxane by monooxygenase-containing bacteria. *Environ. Sci. Technol.* **41**:7330-7336.
3. Mahendra, S., Grostern and L. Alvarez-Cohen. 2012. The impact of common co-contaminants and inducing substrates on degradation and inhibition kinetics of 1,4-dioxane. *Environ. Sci. Technol.*, submitted.

Introduction

The metabolism of dioxane in mammalian systems has been fairly well studied while evaluating the nature of its carcinogenicity. Evidence from animal studies indicates that dioxane biotransformation involves NADH-dependent oxidation by cytochrome P-450 monooxygenase enzymes (170, 177). Several ethers, such as tetrahydrofuran and methyl tert butyl ether, have also been shown to degrade via cytochrome P-450-catalyzed monooxygenation (142, 154). Since bacterial monooxygenase enzymes are similar in structure, function, and reaction mechanisms to mammalian P-450 enzymes, it is logical that these enzymes may also have the ability to degrade dioxane in reactions analogous to their mammalian counterparts. Indeed, the broad specificity of monooxygenases induced by substrates such as methane (106), propane (165), toluene (94), butane (46), and ethene (25) has been widely reported.

Bioremediation may be an attractive option for the treatment of dioxane as it has been demonstrated that dioxane can be degraded by bacteria via both metabolic and cometabolic reactions. While the metabolic and co-metabolic degradation of dioxane has been demonstrated in both bacteria and fungi, little research has been conducted to identify the responsible enzymes involved in the dioxane degradation or to quantify the kinetics of the degradation reactions.

In addition, a complete pathway for dioxane degradation has not yet been determined. In a study of the pathway of bacterial dioxane degradation, Vainberg et al. (161) detected 2-hydroxyethoxyacetic acid (HEAA) as the terminal product of dioxane degradation by *Pseudonocardia* strain ENV478. They suggested that the degradation was carried out by a monooxygenase similar to putative tetrahydrofuran monooxygenase of *Pseudonocardia tetrahydrofuranoxydans* K1 (154). Neither dioxane mineralization nor its assimilation into biomass was observed. Similarly, HEAA has been reported as the major metabolite of dioxane via a cytochrome P450 monooxygenase reaction in mammalian cells (169, 178). An incomplete dioxane degradation pathway has been reported for mammals and aerobic bacteria. In contrast, HEAA was not detected during dioxane degradation by the fungus *Cordyceps sinensis*, although it was suggested that etherases or oxidases were involved in the degradation (101). The only intermediates of dioxane degradation identified in that study were ethylene glycol, glycolic acid, and oxalic acid.

Finally, dioxane was commonly used as a solvent stabilizer for chlorinated solvents such as 1,1,1-trichloroethene (TCA), and historical practices of production, storage, and disposal of

TCA containing dioxane have caused widespread soil and groundwater pollution by solvents and their stabilizers. As a result, TCA and its abiotic breakdown product 1,1-dichloroethene (DCE) are common co-contaminants with dioxane in groundwater aquifers. To date, the effect of chlorinated solvents on dioxane biodegradation has not been evaluated.

The goals of this work were to: evaluate the capabilities of monooxygenase-expressing bacteria for degrading dioxane, measure oxidation kinetics for dioxane degradation, determine the pathway for the complete mineralization of dioxane, and examine the potential inhibitory effects of co-contaminating chlorinated solvents on dioxane degradation.

Results and Discussion

To determine the correlation between monooxygenase expression and the ability to degrade dioxane, we tested a number of known monooxygenase-expressing bacterial isolates for their activity towards dioxane (Table 3.1). Two strains, *Pseudonocardia dioxanivorans* strain CB1190 and *Pseudonocardia benzenivorans* strain B5, were capable of growing with dioxane as the sole carbon and energy source. For both cultures the presence of acetylene (which inhibits some monooxygenase enzymes) inhibited dioxane degradation, and activity was not immediately recovered following acetylene removal. Additionally, dioxane degradation did not occur in the absence of molecular oxygen. Combined, this suggests that the enzymes catalyzing the initial oxidation of dioxane in these two strains are monooxygenases. The cell yields for strains CB1190 and B5 were 0.09 ± 0.002 and 0.03 ± 0.002 mg protein (mg dioxane)⁻¹, respectively. The growth kinetics of both strains were well described by the Monod equation, with the maximum growth rate (k) of strain CB1190 exceeding that of strain B5 by a factor of 10 (Fig. 3.1).

Table 3.1. Bacterial strains tested for the ability to degrade dioxane.

bacterial strain	growth substrate	oxygenase expressed	dioxane degradation
<i>Pseudonocardia dioxanivorans</i> CB1190	dioxane	unknown	yes
<i>Pseudonocardia benzenivorans</i> B5	dioxane	unknown	yes
<i>Pseudonocardia</i> K1	THF	tetrahydrofuran MO	yes
<i>Pseudonocardia</i> K1	toluene	tetrahydrofuran MO	yes
<i>Pseudonocardia sulfidoxydans</i>	THF	unknown	no
<i>Pseudonocardia hydrocarbonoxydans</i>	glucose	unknown	no
<i>Methylosinus trichosporium</i> OB3b	methane	soluble methane MO	yes
<i>Methylosinus trichosporium</i> OB3b	methane	particulate methane MO	no
<i>Mycobacterium vaccae</i> JOB5	propane	propane MO	yes
<i>Rhodococcus</i> RR1	toluene	unknown	yes
<i>Methylibium petroleiphilum</i> PM1	MTBE	unidentified MO	no
<i>Methylibium petroleiphilum</i> PM1	toluene	unidentified MO	no
<i>Burkholderia cepacia</i> G4	toluene	toluene-2-MO	yes
<i>Ralstonia pickettii</i> PKO1	toluene	toluene- <i>p</i> -MO	yes
<i>Pseudomonas mendocina</i> KR1	toluene	toluene-4-MO	yes
<i>Pseudomonas putida</i> mt-2	toluene	toluene-side chain-MO	no
<i>Escherichia coli</i> TG1(T2MO)	LB broth	toluene-2-MO	yes
<i>Escherichia coli</i> TG1(TpMO)	LB broth	toluene- <i>p</i> -MO	yes
<i>Escherichia coli</i> TG1(T4MO)	LB broth	toluene-4-MO	yes
<i>Escherichia coli</i> TG1(TMO)	LB broth	toluene- <i>o</i> -xylene-MO	no
<i>Pseudomonas</i> JS150	toluene	toluene-2,3-DO	no
<i>Pseudomonas putida</i> F1	toluene	toluene-2,3-DO	no
<i>Escherichia coli</i> pCR 2.1-TOPO	LB broth	none	no

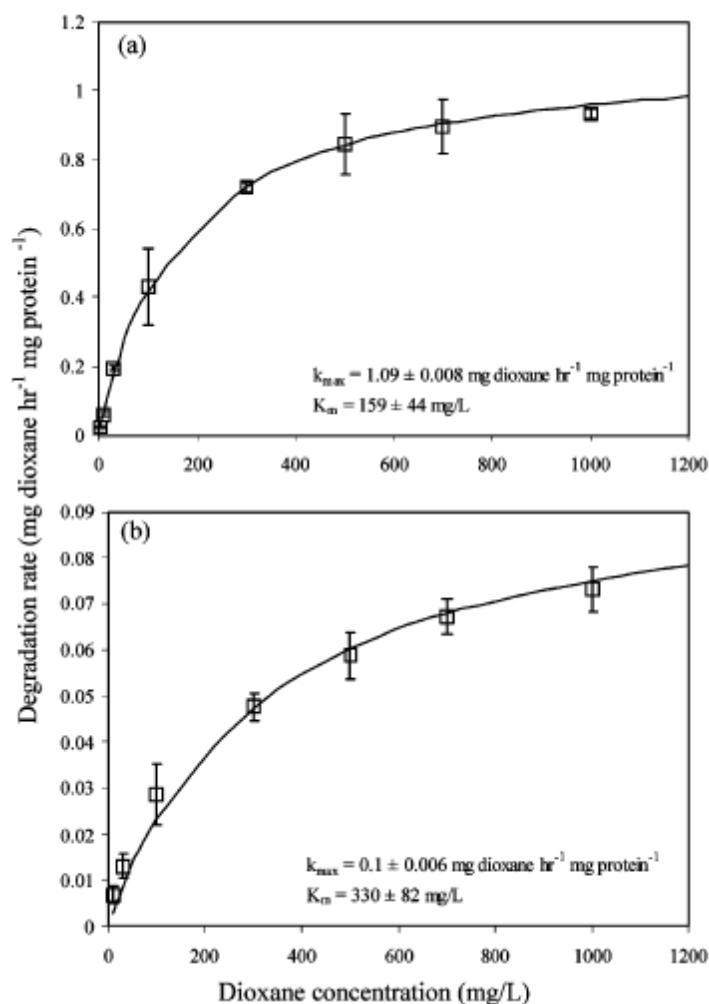


Figure 3.1. Monod plot of dioxane degradation by (a) *Pseudonocardia dioxanivorans* CB1190 and (b) *Pseudonocardia benzenivorans* B5 over a range of concentrations. The parameters k_{max} and K_m were calculated by nonlinear regression minimizing the difference between measured and modeled rates at each concentration. The error bars represent the range of calculated values from each replicate.

Co-metabolic degradation of dioxane was detected in seven bacterial isolates following growth on primary substrates known to induce monooxygenases (Table 3.1). For these strains, inducing substrates leading to dioxane transformation included methane, propane, tetrahydrofuran and toluene. As with the metabolic dioxane degraders, dioxane degradation by co-metabolizing strains was inhibited by acetylene, supporting the hypothesis that a monooxygenase is responsible for dioxane degradation. This role for monooxygenases was further supported by the demonstration that recombinant *E. coli* heterologously expressing monooxygenase genes from known dioxane co-metabolic degraders was able to degrade dioxane, whereas *E. coli* with empty expression vector lacked this activity.

For most of the cultures that degraded dioxane cometabolically, degradation ceased before the dioxane was completely consumed, indicating that the resting cells have a limited transformation capacity for dioxane. This limited transformation capacity may be caused by limited reductant (NADH) to fuel the monooxygenase, toxicity due to dioxane, and/or toxicity due to dioxane metabolites. We examined the reductant limitation issue with *Methylosinus*

trichosporium strain OB3b by supplying dioxane-degrading cell suspensions with formate, which can be used to generate NADH. However, the addition of formate did not have any effect on the rate or extent of dioxane transformation, indicating that reductant limitation did not lead to the cessation of dioxane degradation. Given that OB3b tolerates dioxane concentrations up to 500 mg/L, dioxane toxicity is not likely to cause the observed phenomenon. After limited dioxane degradation OB3b cell suspensions failed to oxidize methane, suggesting that dioxane transformation products exert irreversible toxicity to the monooxygenase enzyme and/or the cells in general.

We next examined the pathway for dioxane degradation and metabolism using metabolic and co-metabolic dioxane-degrading bacterial strains. Cell suspensions were incubated with uniformly ^{14}C labeled dioxane and the distribution of the ^{14}C label in cell biomass, CO_2 , volatile and non-volatile fractions over time was determined. Dioxane-metabolizing strain CB1190 and all four co-metabolic dioxane degraders tested generated $^{14}\text{CO}_2$ and ^{14}C -labeled volatile organic acids, while only CB1190 accumulated ^{14}C label in biomass (Fig. 3.2).

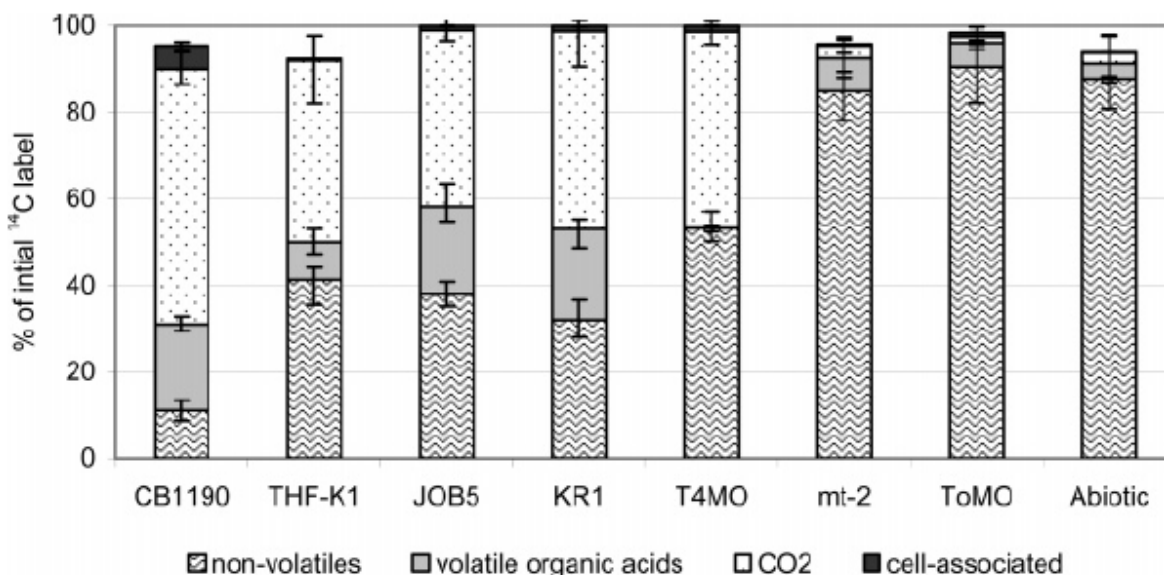


Figure 3.2. Final distribution of ^{14}C derived from the degradation of dioxane after 48 h incubation. Strain metabolizing dioxane: CB1190; strains co-metabolizing dioxane: *Pseudonocardia tetrahydrofuranoxidans* K1, *Mycobacterium vaccae* JOB5, *Pseudomonas mendocina* KR1, *E. coli* heterologously expressing toluene-4-monooxygenase (T4MO), *Pseudomonas putida* mt-2, *E. coli* heterologously expressing toluene/*o*-xylene monooxygenase (ToMO), or sterile abiotic control.

For specific detection of dioxane transformation intermediates, Fourier transform ion cyclotron resonance-mass spectrometry and direct infusion MS-MS were used. The post-hydroxylation, post-ether cleavage metabolite 2-hydroxyethoxyacetic acid (HEAA) was detected transiently in samples from all dioxane-degrading strains. HEAA was subsequently degraded into a mixture of linear dihydroxy-substituted ethoxyacetic acids. Ethylene glycol, glycolic acid, glyoxylic acid, and oxalic acid were also detected.

On the basis of the intermediates identified during dioxane degradation by monooxygenase-expressing bacteria, a complete biodegradation pathway was proposed (Fig. 3.3).

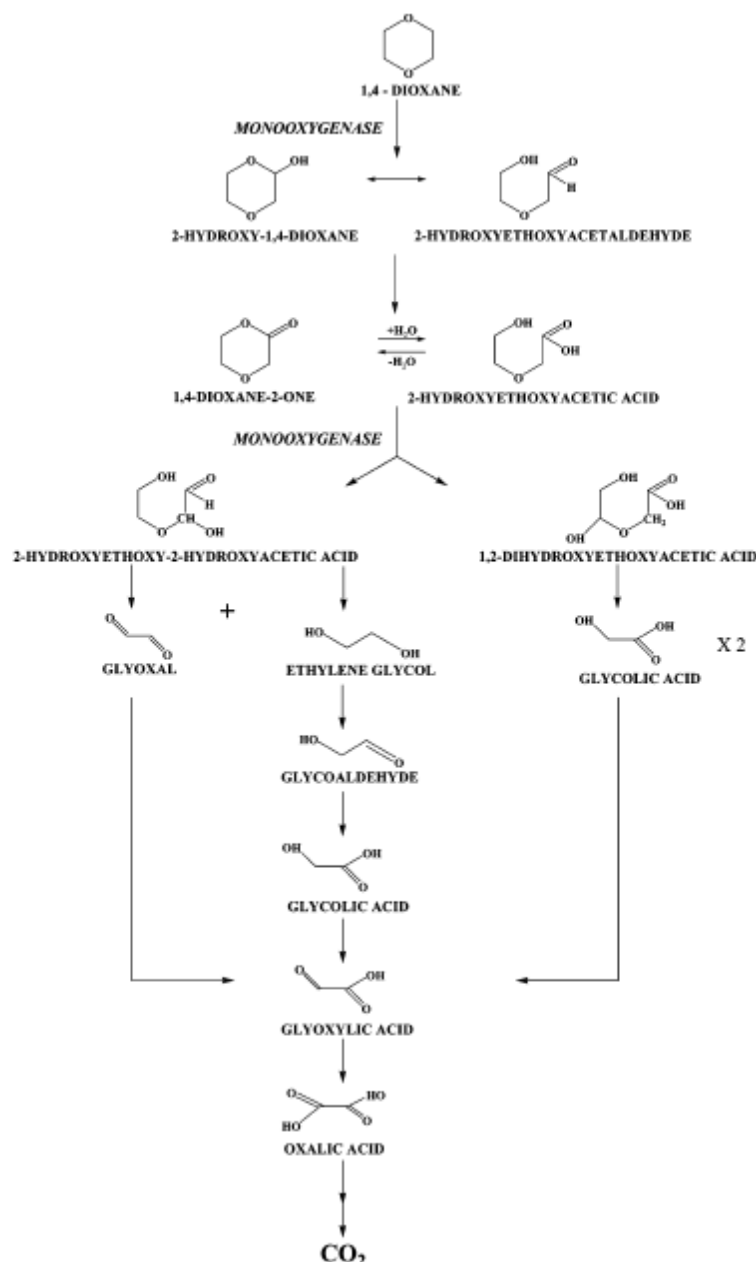


Figure 3.3. Observed biodegradation pathway of dioxane by monooxygenase-expressing bacteria in this study. All degradation products shown in this pathway were identified using a combination of analytical techniques.

The generation of CO₂ from ¹⁴C-labeled dioxane by strains performing co-metabolic dioxane degradation was a surprising result. In a separate study using the dioxane co-metabolic degrader *Pseudonocardia* strain ENV478, Vainberg et al. (161) detected HEAA as the terminal product of dioxane degradation, and mineralization to CO₂ was not found.

Since dioxane contamination often co-occurs in groundwater with chlorinated solvents, particularly 1,1,1-trichloroethane (TCA) and 1,1-dichloroethene (DCE), we explored how these chlorinated co-contaminants affect bacterial dioxane degradation. Strain CB1190 was pre-grown on dioxane, and then dioxane degradation by washed cell suspensions was tested in the presence

of varying concentration of TCA or DCE. Strain CB1190 cells did not degrade either chlorinated solvent. For both TCA and DCE, rates of dioxane transformation decreased inversely with co-contaminant concentration. The observed effects were best modeled as mixed inhibition. After exposure to TCA or DCE and then subsequent washing and resuspension in fresh medium, strain CB1190 cells degraded dioxane at rates similar to those for cells not exposed to chlorinated solvent, indicating that the observed inhibitory effects were reversible.

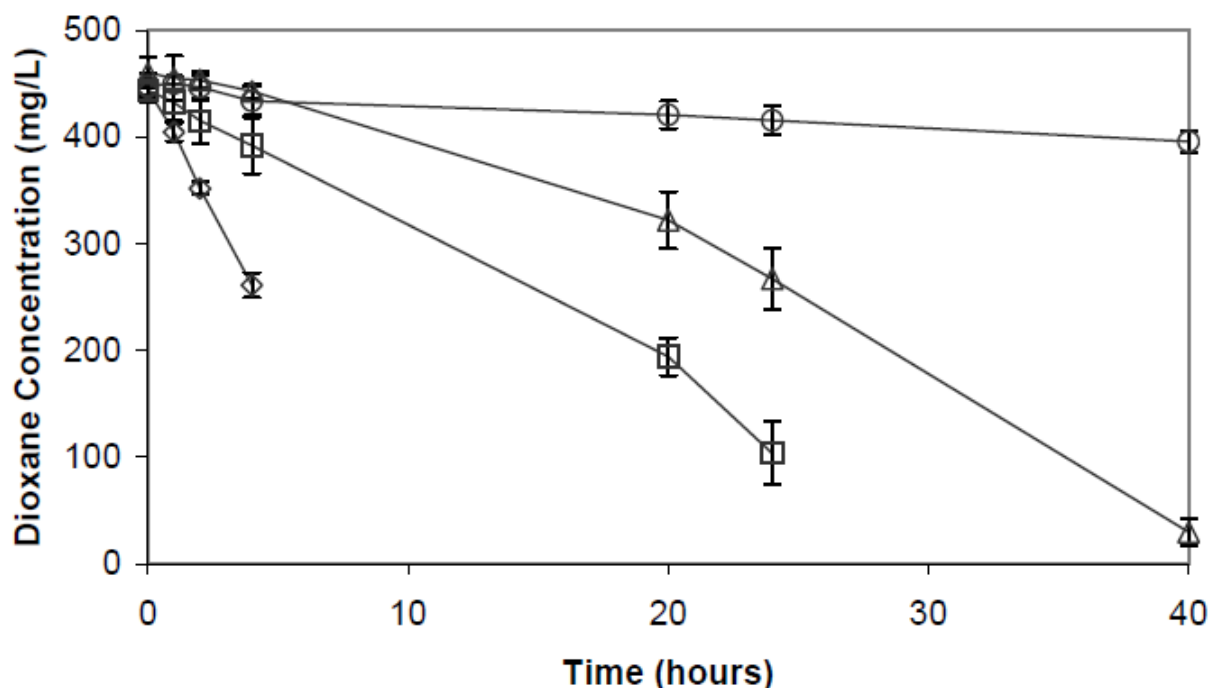


Figure 3.4. Degradation of dioxane by CB1190 in the presence of various concentrations of TCA. Diamonds, 0 mg/L TCA; squares, 0.1 mg/L TCA; triangles, 1 mg/L TCA; circles, 10 mg/L TCA. The error bars represent the range of triplicates.

We next tested if TCA or DCE effect dioxane degradation by *Pseudomonas mendocina* KR1 or *E. coli* heterologously expressing toluene-4-monooxygenase (*E. coli* T4MO). These bacteria degrade dioxane co-metabolically after growth on toluene or LB broth, respectively. The presence of either TCA or DCE negatively affected dioxane degradation by both strains, with the rates of dioxane transformation decreasing inversely with initial co-contaminant concentration (see Fig. 3.4 for strain CB1190 data with TCA). Unlike with strain CB1190, strain KR1 and *E. coli* T4MO degraded TCA and DCE simultaneous with dioxane transformation. Additionally, unlike with strain CB1190, TCA and DCE appeared to irreversibly impair dioxane degradation by strain KR1 and *E. coli* T4MO, since dioxane degradation rates were not restored once the cells were washed and re-suspended in fresh medium. The reason for this irreversible inhibition is unclear.

Conclusion

This work confirmed a role for monooxygenases in dioxane degradation by a number of bacterial strains. Kinetics of dioxane degradation were determined, as was a pathway for the

complete mineralization of dioxane to CO₂ by the dioxane metabolizing bacterium *Pseudonocardia dioxanivorans* strain CB1190. The biodegradation of dioxane via this pathway is unlikely to cause accumulation of toxic byproducts in the environment. Additionally, the inhibition of dioxane degradation by chlorinated solvent co-contaminants indicates that these co-contaminants must be considered when designing bioremediation strategies for removing dioxane from polluted groundwater systems.

4. Sequencing and analysis of the genome of the dioxane degrading actinomycete *Pseudonocardia dioxanivorans* strain CB1190

This is a summary of work published in:

1. Sales, C. M., S. Mahendra, A. Grostern, R. E. Parales, L. A. Goodwin, T. Woyke, M. Nolan, A. Lapidus, O. Chertkov, G. Ovchinnikova, A. Sczyrba, and L. Alvarez-Cohen. 2011. Genome sequence of the 1,4-dioxane-degrading *Pseudonocardia dioxanivorans* strain CB1190. *J. Bacteriol.* 193:4549-4550.
2. Sales, C. M. 2012. Functional genomics of bacterial degradation of the emerging contaminants 1,4-dioxane and N-nitrosodimethylamine (NDMA). Ph.D. thesis. University of California, Berkeley.
3. Additional results presented here are in publications in preparation.

Introduction

The cyclic ether dioxane is increasingly found to contaminate groundwater. The potential toxicity of dioxane has motivated research into how to remove the contaminant from groundwater. One potential method is bioremediation.

Although anaerobic biodegradation of dioxane has been observed with sludge enriched with iron-reducing bacteria (139), most reports of biodegradation of dioxane by fungi and bacteria have occurred under aerobic conditions. Several bacterial and fungal isolates have exhibited the ability to degrade dioxane aerobically (13, 19, 66, 89, 101, 112, 125, 127, 141, 161, 180). However, only five of those isolates are capable of growing on dioxane as a sole carbon and energy source. These include four bacterial strains, *Rhodococcus ruber* 219 (13), *Pseudonocardia dioxanivorans* CB1190 (112), *Pseudonocardia benzenivorans* B5 (89), and *Mycobacterium* sp. PH-06 (66), and one fungal species *Cordyceps sinensis* (101).

Described here is the genome sequence of the Gram-positive actinomycete *P. dioxanivorans* strain CB1190. The genome sequence reported here is the first closed genome sequence in the genus of *Pseudonocardia* that is publicly available, and at 73.12%, is one of the highest G+C content strains ever sequenced. Although 454-pyrosequencing data is publicly available for *Pseudonocardia* sp. P1 (Genbank accession ADUJ000000000), which was isolated from the microbial community associated with colonies of the leaf-cutting ant species *Acromyrmex octospinosus* (8), its assembly is fragmented, with 975 contigs. Members of the *Pseudonocardia* genus have been isolated from soils polluted with aromatic chlorinated compounds (62) and from glacial soils in the McMurdo Dry Valley in Antarctica (120). *Pseudonocardia* species have also been identified as endosymbionts living in the tissue of Chinese medicinal plants (23, 43, 124) and as mutualistic ectosymbionts with leaf-cutting ants and their fungal gardens (20, 119). Strains of *Pseudonocardia* have also exhibited industrial and environmental uses. For instance, strains of *P. autotrophica* produce polyene antifungal antibiotics (82), while *P. tetrahydrofuranoxydans* and *P. chloroethenivorans* degrade the environmental pollutants tetrahydrofuran (THF) and chloroethenes, respectively (62, 68, 83).

P. dioxanivorans strain CB1190 was isolated from a dioxane contaminated industrial sludge in Darlington, South Carolina (112). In addition to growth on dioxane, strain CB1190 can grow on the ethers 1,3-dioxane, THF, tetrahydropyran, 2-methyl-1,3-dioxolane, butyl methyl ether, diethyl ether, and the aromatic compounds benzene and toluene (90, 112). Strain CB1190 also grows autotrophically on hydrogen and carbon dioxide (112), and it can fix nitrogen gas

(90). Even though degradation intermediates have been identified and biochemical assays have suggested the involvement of monooxygenase enzymes (89, 91), the genes that encode for enzymes in the strain CB1190 dioxane metabolic pathways are unknown. The genome sequence described here will allow us to better understand dioxane metabolism, as well as explore the functional capabilities of *P. dioxanivorans* as a toxic pollutant degrader. Furthermore, the genome sequence will provide the first ever insight into the genetics of the genus *Pseudonocardia* in comparison to the genomes of medically and industrially important members of the family *Pseudonocardiaceae*, which include the erythromycin producing *Saccharopolyspora erythraea* NRRL2338 (108), the natural pesticide producing *Saccharopolyspora spinosa* NRRL18395 (111), the rifamycin producing *Amycolatopsis mediterranei* U32 (183) and *Amycolatopsis mediterranei* S699 (164), and the respiratory disease-causing *Saccharomonospora viridis* DSM 43017 (114).

Materials and Methods

Culture conditions

Cells of strain CB1190 used for genomic DNA extraction were grown in ammonium mineral salts (AMS) liquid medium (112). Filter-sterilized dioxane was added to the culture medium and cultures were incubated aerobically while shaking at 30°C.

Genomic DNA extraction

Large quantities of high molecular weight genomic DNA from strain CB1190 were isolated by modifying the CTAB DNA isolation method from the Department of Energy Joint Genome Institute (JGI). Genomic DNA was extracted from cells of strain CB1190 collected by filtration. The cells were scraped from the filter and resuspended in 740 µL of TE buffer (10 mM tris, 1 mM EDTA, pH 8.0) in microcentrifuge tubes. In order to lyse the cells, the cell suspensions were gently mixed with 20 µL of lysozyme (100 mg/mL) and incubated at room temperature for 20 min. Then, 40 µL of 10% SDS (sodium dodecyl sulfate) and 8 µL of Proteinase K (10 mg/mL) were added to each tube and incubated for 12 h at 37°C. Next, 100 µL of 5 M NaCl and 100 µL of CTAB/NaCl heated to 65°C were added to each tube and incubated at 65°C for 10 min. The CTAB/NaCl mixture was prepared by dissolving 4.1 g of NaCl and 10 g of CTAB (hexadecyltrimethyl ammonium bromide) in 100 mL of DI water by constant mixing at 65°C for 3 h and then sterilized by autoclaving. Following incubation with CTAB, the resultant lysate was gently mixed with 0.5 mL of chloroform:isoamyl alcohol (24:1) and then centrifuged for 10 min at maximum speed (21,000 x g) for 10 min at room temperature. The upper aqueous phase was transferred to a clean microcentrifuge tube and gently mixed with 0.5 mL of phenol:chloroform:isoamyl alcohol (25:24:1) (pH 8.0) and then centrifuged for 10 min at maximum speed at room temperature. The upper aqueous phase was transferred to a clean microcentrifuge tube and then mixed with 0.6 volumes of cold isopropanol (-20°C) and incubated at room temperature for 30 min to precipitate the genomic DNA. The genomic DNA was then pelleted by spinning each tube at maximum speed (21,000 x g) for 15 min. The pellets were washed with 500 µL of 70% ethanol. After the supernatant was discarded, the pellet was dried at room temperature for 10 min. Finally, each pellet was resuspended in 20 µL of TE buffer plus RNase (final concentration of 0.1 mg/mL). The size, quantity, and quality of the isolated high molecular weight genomic DNA were determined on a 1% agarose gel.

Genomic sequencing and assembly

The draft genome of *P. dioxanivorans* strain CB1190 was generated at the DOE Joint genome Institute (JGI) using a combination of Illumina (12) and 454 sequencing technologies (92). For this genome we constructed and sequenced an Illumina GAii shotgun library which generated 33,134,275 reads totaling 1,192 Mb, a 454 Titanium standard library which generated 1,211,248 reads and paired-end 454 libraries with an average insert sizes of 3, 10, and 18 kb which generated 210,350 reads for a total of 504.6 Mb of 454 data. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov/>. The initial draft assembly contained 283 contigs in 8 scaffolds. The 454 Titanium standard data and the 454 paired end data were assembled together with Newbler [Version 2.3]. The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data was assembled with VELVET [Version 0.7.63] (182), and the consensus sequence was computationally shredded into 1.5 kb overlapping shreds. We integrated the 454 Newbler consensus shreds, the Illumina VELVET consensus shreds and the read pairs in the 454 paired end library using parallel phrap [Version SPS - 4.24] (High Performance Software, LLC). The software Consed (38) was used in the following finishing process. Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher (48), or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR (J-F Cheng, unpublished) primer walks. A total of 809 additional reactions and 6 shatter libraries were necessary to close gaps and to raise the quality of the finished sequence. The total size of the genome is 7,440,794 bp and the final assembly is based on 285 Mb of 454 draft data which provides an average 38.1x coverage of the genome and 1,011 Mb of Illumina draft data which provides an average 134.8x coverage of the genome. The finished chromosome and plasmid sequences have been deposited to RefSeq and GenBank databases with accession numbers NC_015312-4 and CP002593-8.

Genome annotation and analysis

Annotation of the genome of strain CB1190 was accomplished using the Oak Ridge National Laboratory genome annotation pipeline. Genes were identified using Prodigal (55), followed by a round of manual curation using the JGI GenePRIMP pipeline (113). The predicted coding sequences (CDSs) were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a potential description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE (87), RNAMMer (78), Rfam (42), TMHMM (76), and signalP (10).

Genes involved in transport systems were identified by analyzing putative protein sequences against the Transport Classification Database (<http://www.tcdb.org>) (156), using the basic local alignment search tool for proteins (blastp) (e-value < 1e-5) (4). Signal transduction proteins were found using COG assignments of CDSs and the Microbial Signal Transduction database MiST2.1 (<http://mistdb.com>) (159). Protein domains were identified using the NCBI Conserved Domain Database (CDD) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) or MiST2.1. Genomic islands were identified by SIGI-HMM analysis and visualized using Islandviewer (<http://www.pathogenomics.sfu.ca/islandviewer>) (79).

Results

Genome properties and features of *P. dioxanivorans* strain CB1190

The genome of strain CB1190 has a total size of 7,440,794 bp and consists of four replicons: a circular chromosome (7,096,571 bp), a circular plasmid pPSED01 (192,355 bp), an unclosed circular plasmid pPSED02 (136,805 bp), and a linear plasmid pPSED03 (15,063 bp) (Table 4.1). The unclosed circular plasmid pPSED02 is comprised of three contigs, with sizes of 24,346 bp, 45,552 bp, and 66,907 bp (GenBank Accession numbers CP002595-7), respectively. The average G+C content of the entire genome is 73.12%. The G+C contents of the plasmids are lower than that of the chromosome, with values of 73.41%, 71.15%, 68.38% and 61.83% for the chromosome and plasmids pPSED01, pPSED02, and pPSED03, respectively.

The genome of strain CB1190 contains 6,797 protein-coding sequences (CDS) and 226 pseudo genes. The distribution of CDS, pseudo genes, and hypothetical proteins for each replicon is shown in Table 4.1. The average length of the CDSs across the genome is 963 bp. The chromosome contains 6,495 CDS and 194 pseudo genes while plasmid pPSED01 contains 172 CDS and 20 pseudo genes, plasmid pPSED02 contains 116 CDS and 11 pseudo genes, and plasmid pPSED03 contains 14 CDS and no pseudo genes. Of all predicted proteins in the genome, 4,955 (72.90%) have a putative function ascribed to them while 1,692 genes (26.05%) are annotated as hypothetical proteins. Interestingly, only 3 of the 14 genes (22.4%) on pPSED03 have a putative function ascribed to them.

Table 4.1. Genome feature of *P. dioxanivorans* strain CB1190.

Feature	Genome	Chromosome	Plasmid pPSED01	Plasmid pPSED02	Plasmid pPSED03
Topology		Circular	Circular	Circular	Linear
Length	7,440,794 bp	7,096,571 bp	192,355 bp	136,805 bp	15,603 bp
G+C Content	73.12%	73.41%	71.15%	68.38%	61.83%
Coding Density	87.2%	88.5%	76.1%	80.0%	69.2%
Coding Sequences	6,797	6,495	172	116	14
Pseudo genes	226	194	20	11	0
Average CDS length	963 bp	967 bp	946 bp	851 bp	744 bp
rRNAs	3	3			
tRNAs	47	47			
Hypothetical proteins	1,842	1,692	88	51	11

A total of 59 RNA-encoding sequences were identified in the genome and all are located on the chromosome. Strain CB1190 contains three sets of 16S, 23S, and 5S ribosomal RNA (rRNA) genes, with one set located on the forward strand and two on the reverse strand. Out of the structural RNA sequences, 47 encode transfer RNA (tRNA) genes while the three remaining RNAs encode for a putative catalytic subunit of the RNaseP *rnpB*, a 4.5S RNA component of the signal recognition particle protein translocation system *ffs*, and a small stable RNA *ssrA*, which is involved in a process called *trans*-translation (64, 65).

Comparative genomics of *Pseudonocardiaceae* family

There are currently eleven bacterial strains in the family *Pseudonocardiaceae* for which genome sequence data is publicly available: *Amycolatopsis mediterranei* S699, *Amycolatopsis*

mediterranei U32, *Pseudonocardia dioxanivorans* CB1190 (this study), *Pseudonocardia* sp. P1, *Pseudonocardia* sp. P2, *Saccharomonospora azurea* NA-128, *Saccharomonospora paurometabolica* YIM 9007, *Saccharomonospora viridis* DSM43017, *Saccharopolyspora erythraea* NRRL2338, *Saccharopolyspora spinosa* NRRL18395, and *Thermobispora bispora* DSM43833. However, recent phylogenetic analysis of the 16S rRNA sequences in the genome of *T. bispora* DSM43833 shows that it belongs to the suborder *Streptosporangineae* (Lioliou *et al.*, 2010). Both the suborder *Streptosporangineae* and the family *Pseudonocardiaceae* fall within the order *Actinomycetales*. Further comparative genomic studies, reported herein, perpetuate the concept that *T. bispora* does not belong to the family *Pseudonocardiaceae*. While most members of the family *Pseudonocardiaceae* are Gram-positive, *S. viridis* is a Gram-negative bacterium (114). Of further note, since two groups have sequenced the genome of *S. erythraea* NRRL2338, here the genome sequenced by Oliynyk *et al.* (108) will be referred to as *S. erythraea* NRRL2338-1 and the genome sequenced by the Italian Ministry of Research (Genbank ABFV000000000) as *S. erythraea* NRRL2338-2. The genome of *S. erythraea* NRRL2338-1 is closed, while the genome of *S. erythraea* NRRL2338-2 is comprised of 241 contigs.

Table 4.2 contains a detailed list comparing all genome features of each of the sequenced *Pseudonocardiaceae* strains, including *T. bispora* as an out-group (*Pseudonocardia* sp. P2 is left out because of the lack of annotation data). The genomes of *A. mediterranei* strains were originally sequenced because they are known to produce the antibiotic rifamycin. The two sequenced strains differ by the type of rifamycin that they produce: *A. mediterranei* S699 produces rifamycin B, while strain U32 produces rifamycin SV (164). The complete genome of *A. mediterranei* S699 is the largest in the family *Pseudonocardiaceae*, consisting of a circular chromosome with a total length of 10,236,779 bp. The genome of *A. mediterranei* S699 was assembled by mapping sequencing reads to the genome of *A. mediterranei* U32 as a reference (164). The genome of *A. mediterranei* U32 also contains a single circular chromosome. The smallest of the sequenced genomes in the family *Pseudonocardiaceae* is that of *S. viridis*, with a total length of 4,308,349 bp. Although only draft sequences are available, the other two *Saccharomonospora* species have similar genome sizes (4,770,125 bp and 4,592,308 bp, respectively). The two sequences for the genome of *S. erythraea* NRRL2338 vary because the genome of *S. erythraea* NRRL2338-1 is closed, while the genome of *S. erythraea* NRRL2338-2 is a draft. As shown in Table 4.2, the draft genome of *S. spinosa* is larger than the genome sequences for *S. erythraea*. In addition to strain CB1190, the only *Pseudonocardia* species with genome sequencing data are *Pseudonocardia* sp. P1 and *Pseudonocardia* sp. P2; however, their genomes are poorly assembled. Whole genome shotgun sequencing data is available for *Pseudonocardia* sp. P2 (GenBank AEGE000000000), but it has not been annotated and contains an extremely large number of contigs. Currently, the genome of *Pseudonocardia* sp. P2 has been assembled into 1,778 contigs consisting of 8.7 Mb.

Table 4.2. Features of sequenced *Pseudonocardia* genomes^a

Features	<i>P. dioxanivorans</i>	<i>A. mediterranei</i> S699	<i>A. mediterranei</i> U32	<i>Pseudonocardia</i> sp. P1	<i>S. azurea</i>	<i>S. paurometabolica</i>	<i>S. viridis</i>	<i>S. erythraea</i> NRRL2338-1	<i>S. erythraea</i> NRRL2338-2	<i>S. spinosa</i>	<i>T. bisporea</i>
Status	Finished	Finished	Finished	Draft	Draft	Draft	Finished	Finished	Draft	Draft	Finished
Topology	Circular	Circular	Circular	Unknown	Unknown	Unknown	Circular	Circular	Unknown	Unknown	Circular
Genome size	7,096,571 bp	10,236,779 bp	10,236,715 bp	6,388,771 bp	4,770,125 bp	4,592,308 bp	4,308,349 bp	8,212,805 bp	8,079,083 bp	8,527,776 bp	4,189,976 bp
GC content	73%	71%	71%	73%	70%	71%	67%	71%	71%	68%	72%
Coding Density	88%	90%	89%	88%	90%	89%	86%	85%	83%	80%	83%
CDSs ^c	6,797	9,575	9,228	6,620	4,543	4,876	3,828	7,197	7,305	8,215	3,546
Pseudo genes	226	ND ^d	ND	ND	ND	ND	78	1	ND	ND	50
Hypothetical genes	1,842 (27.1%)	3,173 (33.1%)	2,854 (30.9%)	2,206 (33.3%)	1,127 (24.9%)	1,271 (26.0%)	980 (25.6%)	2,204 (30.6%)	2,326 (31.8%)	3,339 (40.6%)	980 (27.6%)
Plasmids ^b	3, free replicons	1, integrated	2, integrated	Unknown	Unknown	Unknown	0	4, integrated	Unknown	Unknown	0
rRNA operons	3	4	4	1	1	1	4	4	1	1	3
tRNA genes	47	52	52	50	48	47	53	50	50	50	53
Contigs	NA ^e	NA	NA	875	187	878	NA	NA	241	118	NA

All information is from GenBank and Refseq databases (*P. dioxanivorans*, NC_015312-4 and CP002593-8; *A. mediterranei* S699, CP002896; *A. mediterranei* U32, NC_014318; *Pseudonocardia* sp. P1, NZ_ADUJ000000000; *S. azurea*, NZ_AGIU000000000; *S. paurometabolica*, NZ_AGIT000000000; *S. viridis*, NC_013159; *S. erythraea* NRRL2338-1, NC_009142; *S. erythraea* NRRL2338-2, NZ_ABFV000000000; *S. spinosa*, NZ-AEYC000000000; *T. bisporea*, NC_014165).

^a *T. bisporea* does not belong in the family *Pseudonocardiaceae*, it belongs to the suborder *Streptosporangineae*

^b Information regarding integrated plasmids in *S. erythraea* NRRL2338-1 and *A. mediterranei* U32 are from te Poele *et al.* (152) and Zhao *et al.* (183), respectively.

^c CDS – Protein coding sequence

^d ND – Not determined

^e NA – Not applicable

Calculation of the number of reciprocal best protein blast hits between each pair of genomes (Table 4.3) reflects the phylogeny, classification, genome sizes, and completeness of the genome assembly of the strains within the family of *Pseudonocardiaceae*. Using a blastp cutoff e-value of 1e-5 and a minimum coverage of 70% to determine reciprocal blast hits between a pair of genomes, strain CB1190 is shown to share the most CDSs with the other *Pseudonocardia* species. Even though *Pseudonocardia* sp. P1 has 1,518 more predicted protein encoding genes than *Pseudonocardia* sp. P2, they have nearly identical numbers of reciprocal blast hits with strain CB1190 (3,251 and 3,242, respectively). Among all three *Pseudonocardia* species, *Pseudonocardia* sp. P1 and *Pseudonocardia* sp. P2 are more closely related to each other, sharing 4,210 common proteins. Among other genera, *Pseudonocardia* genomes are most related to *Amycolatopsis* genomes, with 39-48% similarity. In agreement with its 16S rRNA phylogenetic relationship, *T. bisporea* shares the least number of protein-coding genes with members of the family *Pseudonocardiaceae* (1,840 to 2,136). As expected, a high degree of similarity is seen between genomes from the same species (95% -99% between *A. mediterranei*). Surprisingly, the two genomes from the same strain of *S. erythraea* appear to be less similar (93% -96%), but this may be due to the fact that the genome of *S. erythraea* NRRL2338-2 is unfinished while the genome of *S. erythraea* NRRL2338-1 is closed.

Mobile genetic elements

The strain CB1190 genome contains a large number of mobile genetic elements (MGEs) located on the chromosome and two of the plasmids, pPSED01 and pPSED02. A total of 129 open reading frames (ORFs) encode for transposases and integrases, belonging to the insertion sequence families of IS3, IS4, IS21, IS111A and IS116. The majority of MGEs, 106 of them, are located on the chromosome, while 13 and 9 of them are located on plasmids pPSED01 and pPSED02, respectively.

Using SIGI-HMM, a number of genomic islands (GIs) were predicted on the chromosome of strain CB1190. Two clusters of predicted GIs were located in the regions of 1.30-1.38 Mbp (Psed_1228 to Psed_1324) and 3.68-3.74 bp (Psed_3442 to Psed_3501). These regions have a G+C content of 68.7% and 68.3%, respectively, which is lower than the G+C content of the chromosome (73.41%). Both regions contain a number of transposase and integrase genes, so they are likely a result of a horizontal gene transfer event. The first cluster contains a number of genes encoded for heavy metal resistance of mercuric compounds. The second cluster contains a number of transporter and chaperone genes that may help the cell to overcome toxicity. The inclusion of a plasmid partitioning gene *parB* (Psed_3493) in the second cluster suggests that this GI originated from a plasmid.

Signal transduction systems

A total of 657 ORFs were determined to be signal transduction proteins. These signal transduction proteins allow cells to respond to environmental changes. According to the Microbial Signal Transduction Database (MiST2.1) analysis, 51 ORFs were predicted to encode putative histidine kinases and 57 were predicted to encode putative response regulators. Based on co-localization, 37 two-component regulatory systems were identified. These two-component systems consist of a histidine kinase (HK) and a response regulator (RR). The HK senses specific environmental stimuli, while the RR controls the cellular response by the differential expression of certain genes.

Table 4.3. Reciprocal best blast hits between *Pseudonocardiaceae*.

The values indicate the number of shared proteins between a pair of genomes, based on a count of all reciprocal best blastp hits between the pairs, using a minimum e-value = 1e-5 and a minimum coverage = 70%.

	<i>P. dioxanivorans</i> CB1190	<i>A. mediterranei</i> S699	<i>A. mediterranei</i> U32	<i>Pseudonocardia</i> sp. P1	<i>Pseudonocardia</i> sp. P2	<i>S. azurea</i>	<i>S. paurometabolica</i>	<i>S. viridis</i>	<i>S. erythraea</i> NRRL2338-1	<i>S. erythraea</i> NRRL2338-2	<i>S. spinosa</i>	<i>T. bispora</i>
<i>P. dioxanivorans</i> CB1190	-	3219	3209	3251	3242	2392	2368	2210	2842	2845	2789	1890
<i>A. mediterranei</i> S699	3219	-		3035	3138	3109	2943	2846	3666	3665	3358	2136
<i>A. mediterranei</i> U32	3209	9137	-	3033	3135	3106	2939	2846	3657	3657	3346	2135
<i>Pseudonocardia</i> sp. P1	3251	3035	3033	-	4210	2408	2387	2241	2989	2990	2815	1840
<i>Pseudonocardia</i> sp. P2	3242	3138	3135	4210	-	2399	2387	2252	2950	2936	2813	1872
<i>S. azurea</i>	2392	3109	3106	2408	2399	-	3043	3195	2901	2893	2660	1849
<i>S. paurometabolica</i>	2368	2943	2939	2387	2387	3043	-	2849	2738	2719	2568	1807
<i>S. viridis</i>	2210	2846	2846	2241	2252	3195	2849	-	2662	2656	2472	1785
<i>S. erythraea</i> NRRL2338-1	2842	3666	3657	2989	2950	2901	2738	2662	-	6901	4091	2024
<i>S. erythraea</i> NRRL2338-2	2845	3665	3657	2990	2936	2893	2719	2656	6901	-	4063	2022
<i>S. spinosa</i>	2789	3358	3346	2815	2813	2660	2568	2472	4091	4063	-	1955
<i>T. bispora</i>	1890	2136	2135	1840	1872	1849	1807	1785	2024	2022	1955	-

As is common in bacteria, one-component signal transduction systems are more abundant in strain CB1190 than two-component systems. One-component systems consist of a single protein that contains input (sensory) and output (regulatory) domains but lack phosphotransfer domains (*i.e.*, receiver (RR) and transmitter (HK) domains) typical of two-component systems (158). A total of 502 genes encoding one-component signal transduction proteins were identified that contain at least one conserved output domain. The output domains in one-component systems are significantly more diverse than those in two components, containing mostly DNA-binding domains but also RNA-binding domains, phosphohydrolase (HD) domains, phosphatase domains, protein kinase (Pkinase) domains, and nucleotide (adenylate and di-guanylate) cyclase domains. Even among the DNA-binding domains, there is a higher degree of variety in one-component systems than two-component systems.

A third important class of bacterial signal transduction proteins are extracytoplasmic function (ECF) σ factors. All bacteria have a primary σ factor that controls basal level expression of most genes. Bacterial σ factors assist RNA polymerase in initiating transcription by allowing

them to bind to specific gene promoters (52). Alternative σ factors, such as ECF σ factors, are activated in the presence of certain environmental stimuli (51). In the absence of a stimulus, alternative σ -factors are kept inactive through protein-protein binding with anti- σ factors (18). A total of 50 ORFs encode for putative RNA polymerase σ factors and 34 of these contain conserved ECF σ factor domains recognized by MiST2.1. Despite being unrecognized by MiST2.1, the remaining σ factors were found to have output domains homologous to the common σ^{70} factor domains.

Among *Pseudonocardiaceae*, strain CB1190 has fewer signal transduction proteins than *A. mediterranei* U32 (1,141), *S. erythraea* NRRL2338-1 (738), and *S. erythraea* NRRL2338-2 (727), but more than *S. viridis* P101 (323) and *Pseudonocardia* sp. P1 (585). According to the MiST2.1 database, one-component systems in all *Pseudonocardiaceae* make up approximately 70-80% of all signal transduction proteins. In addition, 84 proteins with conserved ECF σ factors in *A. mediterranei* U32, 32 in *S. erythraea* NRRL2338-1, 17 in *Pseudonocardia* sp. P1, and 13 in *S. viridis* were discovered.

Transport systems

The strain CB1190 genome contains many genes associated with transport. A total of 961 genes (13.7% of all CDS) encoding a broad range of transport functions were identified according to blastp searches (e-value < 10e-5) against the Transport Classification Database. The majority of the predicted transport genes were located on the chromosome, while only 21, 15, and 1 were identified on the plasmids pPSED1, pPSED2, and pPSED3, respectively. Twenty six genes were predicted to encode for transmembrane channels, including those involved in the transport of potassium, ammonia, urea-amide, and magnesium-cobalt-nickel. A total of 230 genes were homologous to carriers that include specificity for sugars, aromatic compounds, drugs, heavy metals, nitrate, nitrite, sulfate, amino acids, and nucleosides. Ninety one of these potential carriers were members of the Major Facilitator Superfamily (MFS). The largest group of transporters was the ATP-binding Cassette (ABC) Superfamily, consisting of 283 genes or 29.4% of all transport associated genes in the strain CB1190 genome. These ABC transporters were predicted to be involved in P-P hydrolysis driven transport of substrates such as carbohydrates, heavy metals, drugs, liposaccharides, lipoproteins, nitrate, nitrite, bicarbonate, organic acids, cholesterol, amino acids, urea, and oligopeptides.

A significant number of ABC (44) and MFS (45) transporters were predicted to be involved in drug-resistance or export of drugs. This accounts for 15% of ABC transporters and 35% of MFS transporters. Sixteen of the ABC transporter genes were found to be members of the Drug Exporter-1 (DrugE1) Family. Three of the MFS transporter genes (Psed_0339, Psed_4022, and Psed_6185) were identified as members of the Drug:H⁺ Antiporter-1 (DHA1) Family, while 35% of all MFS transporter genes were members of the Drug:H⁺ Antiporter-2 (DHA2) Family.

The strain CB1190 genome contains a number of genes involved in heavy metal resistance. A cluster of genes (Psed_6718 to Psed_6725) is homologous to the mercury resistance operon of *Streptomyces lividans* (134). The genome of strain CB1190 also contains two alkylmercury lyases (Psed_1313 and Psed_1314) and a mercuric reductase (Psed_1317), which are involved in the detoxification of mercury. An arsenate resistance operon, containing genes for a putative arsenate reductase (Psed_2407), an arsenical resistance protein (Psed_2408), and an arsenic-related regulatory protein ArsR (Psed_2409), was also identified.

Secretion systems

Genes for the general secretory pathway are present in the strain CB1190 genome. Unlike *E. coli* but similar to *Comamonas testosteroni* CNB-2 (88), the Sec genes identified in strain CB1190 were not clustered, except for *secF*, *secD*, and *yajC* (Psed_3580 to Psed_3582), which encode a putative protein complex. The strain CB1190 genome also contains genes predicted to be members of the Type III (IIISP) and Type IV (IVSP) secretory pathways, which are typically involved in the transport of virulence factors. Three type IV secretory pathway genes are located on plasmid pPSED2 (Psed_6900, Psed_6901, and Psed_6994).

Chaperone proteins

The genome of strain CB1190 contains 51 chaperone genes. Two genes (Psed_6730 and Psed_6374) were annotated as *hypC/hupF*, chaperones involved in the assembly of hydrogenases. A gene annotated as a nitrite reductase molybdenum cofactor assembly chaperone (Psed_4302) was also identified. A number of genes were also identified to encoded for heat shock and cold shock proteins. Similar to *Rhodococcus jostii* RHA1 (137) and *Gordonia* sp. TY-5 (72), a chaperonin *GroEL* (Psed_0633) is found within a cluster of genes that encode for a putative propane monooxygenase.

Nitrogen metabolism

Strain CB1190 can use multiple nitrogen sources (90), and this is reflected in the genome annotation. Ammonia in AMS medium can be incorporated by homologues of glutamine synthetase (Psed_2546, Psed_3968, Psed_4574, and Psed_4917), which condenses ammonia with glutamate to form glutamine. Strain CB1190 can use nitrate, and a total of eight ORFs encode nitrate reductases, which convert nitrate to nitrite. One of these nitrate reductase genes (Psed_1088) is in a cluster of genes that encode for subunits of a nitrite reductase (Psed_1089 to Psed_1091), a nitrite transporter (Psed_1087), and an uroporphyrinogen-III synthase HEM4 (Psed_1092). Based on blastp analyses, the reductases in this gene cluster are most similar to assimilatory NAD(P)H-dependent nitrate and nitrite reductases in *Amycolatopsis mediterranei* U32, which function to reduce nitrate to NH_3 via NO_2^- and hydroxylamine. In strain CB1190, the uroporphyrinogen-III synthase catalyzes the synthesis of uroporphyrinogen-III, a precursor for siroheme, which is a cofactor for nitrite reductase (135). In addition to being able to assimilate NO_3^- , strain CB1190 also appears capable of dissimilatory nitrate reduction. The presence of a cluster of four ORFs encoding for subunits of a respiratory nitrate reductase (Psed_4300 to Psed_4303) suggests the ability to use NO_3^- as a terminal electron acceptor under anaerobic conditions. However, genes for respiration of NO_2^- are absent, such as *nrfA*-encoded cytochrome c nitrite reductase, which reduces NO_2^- to ammonia, or the *nirS*-encoded cytochrome cd1 nitrite reductase or the *nirK*-encoded copper nitrite reductase, which can each reduce NO_2^- to nitric oxide (NO). Rather, the NO_2^- is reduced to NH_4^+ by either assimilatory or dissimilatory mechanisms without producing a proton-motive force (*i.e.*, during NO_2^- respiration). Furthermore, nitric oxide and nitrous oxide reductases are missing in the genome, meaning NO_3^- cannot be reduced all the way to dinitrogen (N_2). However, genes are present for nitric oxide dioxygenases (Psed_4325, Psed_5584, and Psed_6551), which may allow strain CB1190 to cope with nitric oxide stress by oxidizing it to nitrate.

It was previously shown that strain CB1190 can fix dinitrogen (N_2) (90), an activity that has not been reported for other *Pseudonocardia*. Evidence for this activity included the conversion of acetylene to ethylene (a common activity of nitrogenases) and the production of

ammonia by cells growing in mineral media void of nitrogen sources (*i.e.*, NH_3 , NO_3^- , or NO_2^-). However, no homologues to nitrogenase genes were found in the genome of strain CB1190, except for two genes annotated as *nifU*-like proteins (Psed_3394 and Psed_6367) and one as a *NifC*-like porter (Psed_5656). Many organisms which are unable to fix dinitrogen contain *nifU*-like proteins because they can assist in the formation of metallocusters, typically FeS clusters, for proteins other than nitrogenases. Similarly, *nifC*-like porters, which are involved in the transport of molybdenum for the molybdenum-containing nitrogenase enzymes, could transport molybdenum for other enzymes or co-factors. While a unique system called superoxide-dependent N_2 (sdn) fixation was described in *Streptomyces thermoautotrophicus*, in which N_2 reduction to ammonium is coupled to oxidation of superoxide produced from O_2 and CODH, the short N-terminal polypeptide sequences for the CODH, the superoxide oxidoreductase, and the novel oxygen-insensitive nitrogenase from *S. thermoautotrophicus* are not sufficient for homology searches (126). Although Hoffmann-Findeklee *et al.* (2002) describe cloning and sequencing of *sdn* genes from *S. thermoautotrophicus* and suggest that the nitrogenase belongs to the molybdenum-hydroxylase family, neither nucleotide nor amino acid sequences are publicly available. Therefore, the N_2 fixation system used by strain CB1190 remains unknown.

General carbon metabolism

The ability of strain CB1190 to grow on mineral medium indicates it can synthesize all of its cellular components, including fatty acids, nucleotides and amino acids, and *in silico* analysis of the genome supports this notion. Complete pathways are present for the synthesis of 22 amino acids, including the selenoamino acids: selenocysteine and selenomethionine. Both the glycolysis and pentose phosphate (PP) pathway are complete, suggesting that glucose can be metabolized through at least two pathways. In addition, the entire citric acid or tricarboxylic acid (TCA) cycle is present.

The genome sequence also suggests that strain CB1190 can metabolize xylose using the isomerase pathway. The genome contains 11 putative xylose isomerase genes (Psed_0624, Psed_1104, Psed_2264, Psed_2595, Psed_3711, Psed_3726, Psed_5635, Psed_6921, Psed_6931, Psed_6943, and Psed_6944), which encode proteins to convert D-xylose into D-xylulose. Also present are genes for two xylulokinases (Psed_1102 and Psed_3672) that can phosphorylate D-xylulose into D-xylulose-5-phosphate, an intermediate in the PP pathway.

Carbon-fixation pathways

Strain CB1190 grows autotrophically by fixing CO_2 in the presence of H_2 (112). Six primary CO_2 fixation pathways have been identified in prokaryotes: 1) the Calvin-Benson-Bassham (CBB) pathway, 2) the Reductive TCA cycle or reverse citric acid cycle, 3) the reductive acetyl Co-A pathway or Wood-Ljungdahl pathway, 4) the 3-hydroxypropionate pathway/methyl-CoA pathway (3-HP), 5) the 3-hydroxypropionate/4-hydroxybutyrate cycle, and 6) the dicarboxylate/4-hydroxybutyrate cycle (77). A search of the strain CB1190 genome for genes related to each of these pathways was conducted, and revealed the presence of a near-complete CBB pathway. Genes potentially contributing to the other five pathways were also identified, but the CB1190 genome lacks homologous enzymes catalyzing several steps in each of these five pathways, so the potential activity of these pathways could not be inferred from genome sequence information.

Monooxygenases

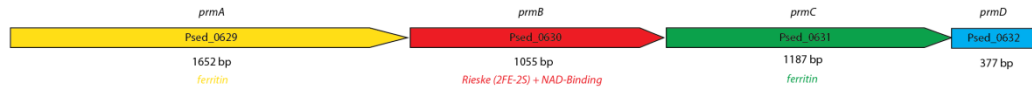
The first step in dioxane biodegradation by strain CB1190 is catalyzed by a monooxygenase reaction ((89, 91) and Chapter 6). Monooxygenases are enzymes that use the energy from NAD(P)H to reduce dioxygen (O_2) into water and a hydroxyl group, which is subsequently incorporated onto a substrate. A total of 84 ORFs are predicted to encode for monooxygenase components. Since monooxygenases catalyze a hydroxylation reaction, they are often referred to and contain proteins encoded as hydroxylases. The genome of strain CB1190 contains 14 ORFs annotated as hydroxylases. In addition, 11 ORFs are predicted to encode for cytochrome P450 (CYP) enzymes, which are known to catalyze monooxygenase reactions of a variety of organic compounds.

Eight gene clusters of bacterial multicomponent monooxygenases (BMMs) are present in the strain CB1190 genome. These BMMs contain two to six co-located genes, including at least a hydroxylase protein and a reductase protein. The reductase protein supplies electrons to the active site of the hydroxylase component by oxidizing NAD(P)H. These electrons allow the hydroxylase component to catalyze the hydroxylation of the substrate. The hydroxylase component of BMMs may consist of two proteins: a large α subunit and smaller β subunit. Classification of BMMs is based upon their substrate specificity. The subunit composition and arrangement of coding sequences differs for each type of BMM. Therefore, the order of co-localized genes and their similarity to sequences of known BMM components were used to categorize and deduce the function of the eight BMM clusters in strain CB1190 (see Figure 4.1).

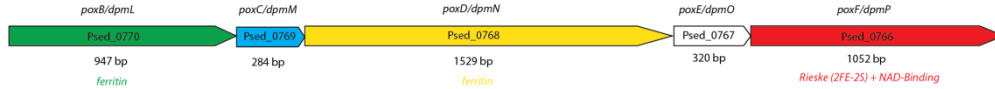
Seven of the BMM clusters are located on the chromosome, including a propane monooxygenase (*prm*) gene cluster (Psed_0629 to Psed_0632) and a putative phenol hydroxylase (Psed_0770 to Psed_0766). Four chromosome-encoded BMMs are similar to aromatic monooxygenases that target benzene, toluene, ethylbenzene, and xylene. The clusters denoted as aromatic monooxygenase 1 (Psed_0815 to Psed_0810) and aromatic monooxygenase 3 (Psed_1436 to Psed_1441) in Figure 4.1 have identical *tmoABDEC* gene structures. Furthermore, the corresponding proteins in aromatic monooxygenase 1 and aromatic monooxygenase 3 share 46.6% to 81.6% identity. Similarly, the order of the genes in aromatic monooxygenase 2 (Psed_1155 to Psed_1159) and aromatic monooxygenase 4 (Psed_6062 to Psed_6058) are identical to one another and are similar to that of the toluene/o-xylene monooxygenase *touABCDEF* gene cluster found in *Pseudomonas stutzeri* OX1 (14) and the toluene 4-monooxygenase *tmoABCDEF* gene cluster found in *Pseudomonas mendocina* KR1 (174, 175). The protein sequences in aromatic monooxygenase 2 and 4 are 37.8% to 50.0% identical.

The simplest BMM found is a two-component 4-hydroxyphenylacetate 3-monooxygenase encoded by Psed_6066-67. This monooxygenase is a flavin adenine dinucleotide (FAD)-dependent monooxygenase. The large component HpaB (Psed_6066) utilizes $FADH_2$ to oxidize 4-hydroxyphenylacetate (HPA) (122). The $FADH_2$ required for hydroxylation is produced by the small subunit *hpaC* (Psed_6066), a flavin:NAD(P)H reductase, which oxidizes NAD(P)H in order to reduce FAD (40).

A. Propane monooxygenase



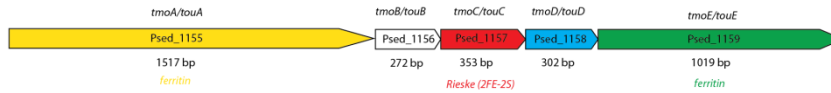
B. Phenol monooxygenase



C. Aromatic monooxygenase 1



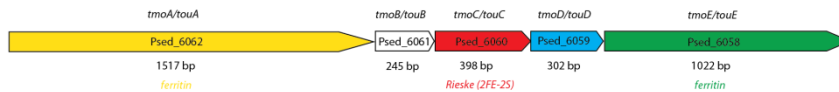
D. Aromatic monooxygenase 2



E. Aromatic monooxygenase 3



F. Aromatic monooxygenase 4



G. 4-Hydroxyphenylacetate monooxygenase



H. Tetrahydrofuran monooxygenase

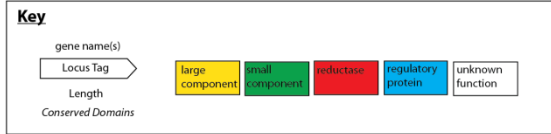
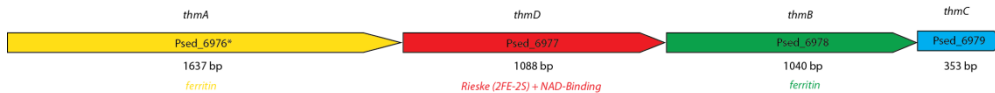


Figure 4.1. Bacterial multicomponent monooxygenases (BMMs) in strain CB1190. Seven of the BMMs (A-G) are located on the chromosome, while the tetrahydrofuran monooxygenase (THFMO) is located on the plasmid pPSED02. *prmA*, propane monooxygenase (PrMO) hydroxylase large subunit; *prmB*, PrMO reductase; *prmC*, PrMO hydroxylase small subunit; *prmD*, PrMO coupling protein; *poxB/dpmL*, phenol hydroxylase P1 protein; *poxC/dpmM*, phenol hydroxylase P2 protein; *poxD/dpmN*, phenol hydroxylase P3 protein; *poxE/dpmO*, phenol hydroxylase P4 protein; *poxF/dmpP*, phenol hydroxylase P5 protein; *tmoA/touA*, toluene monooxygenase (TMO) hydroxylase α subunit; *tmoB/touB*, TMO hydroxylase β subunit; *tmoC/touC*, TMO effector protein; *tmoD/touD*, TMO protein D; *tmoE/touE*, TMO protein E; *HpaB*, 4-hydroxyphenylacetate-3-hydroxylase reductase component, *HpaC*, 4-hydroxyphenylacetate-3-hydroxylase oxygenase component; *thmA*, THFMO oxygenase component α subunit; *thmB*, THFMO oxygenase component β subunit; *thmC*, THFMO coupling protein; *thmD*, THFMO reductase component. Note Psed_6976* is annotated as a pseudogene (GenBank CP002597).

The only non-chromosomal BMM is a tetrahydrofuran (THF) monooxygenase (Psed_6976 to Psed_6979), which is located on plasmid pPSED02 and is in a cluster of nine genes (Psed_6974 to Psed_6982). Eight of the genes are highly homologous to tetrahydrofuran monooxygenase (*thm*) gene clusters in *Pseudonocardia* sp. ENV478, *P. tetrahydrofuranoxydans* K1, and *Rhodococcus* sp. YYL (Genbank HQ99619.1, AJ296087.1, and EU732588.2, respectively). The eight common *thm* genes are *orfY*, *sad*, *thmA*, *thmD*, *thmB*, *thmC*, *orfZ*, and *aldH*. The one gene in this cluster found only in strain CB1190 encodes for a NRAMP family Mn^{2+}/Fe^{2+} transporter (Psed_6982). The existence of this Fe^{2+} transporter makes sense because of the high similarity of THF monooxygenases to binuclear-iron-containing multicomponent monooxygenases (153, 154). Sequence analysis indicates that the *thm* cluster in strain CB1190, excluding the transporter gene, shares 96%, 94%, and 87% identity at the nucleotide level to *Pseudonocardia* sp. ENV478, *P. tetrahydrofuranoxydans*, and *Rhodococcus* sp. YYL, respectively. The discovery of the *thm* cluster on plasmid pPSED02 in strain CB1190 is consistent with previous findings that the *thm* gene sequences in *P. tetrahydrofuranoxydans* had a characteristically low G+C content and was shown to be located on a large plasmid (pPSK50) (154). The *thm* proteins encoded on these plasmids are nearly identical, sharing 91.4 to 99.6% similarity (Fig. 4.2). Furthermore, the GC content of the *thm* cluster in strain CB1190 is 61.3%, while the GC content of the entire plasmid pPSED2 is 68.4%.

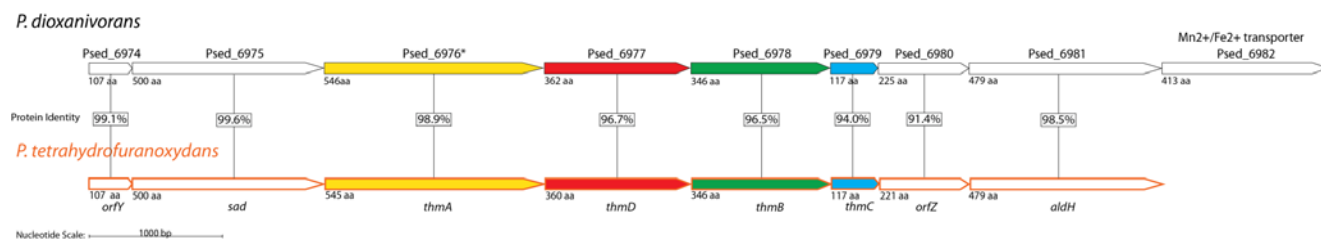


Figure 4.2. THF MO gene clusters from strain CB1190 and *P. tetrahydrofuranoxydans*. *orfY*, unknown function; *sad*, succinate semialdehyde dehydrogenase; *thmA*, THF MO α subunit; *thmB*, THFMO β -subunit; *thmC*, THFMO reductase; *thmD*, THFMO coupling protein; *orfZ*, unknown function; *aldH*, aldehyde dehydrogenase. Yellow indicates monooxygenase large subunit, red indicates reductase, green indicates monooxygenase small subunit, and blue indicates monooxygenase coupling protein.

Dioxygenases

The strain CB1190 genome contains 92 genes annotated as putative dioxygenase components that catalyze the incorporation of both atoms of dioxygen into substrates as hydroxyl groups. Many of these dioxygenases appear to be involved in the degradation of aromatic compounds. Two of the dioxygenase genes are located on plasmid pPSED02: an extradiol ring-cleave dioxygenase class III protein subunit B (Psed_6935) and a phthalate 3,4-dioxygenase ferredoxin subunit (Psed_6938). Half of the putative dioxygenases (46) are annotated as glyoxylase/bleomycin resistance protein/dioxygenases and are distributed throughout the chromosome. Some of the dioxygenases appear to be involved in the metabolism of aromatic amino acids, such as phenylalanine with 3-carboxyethylcatechol 2,3-dioxygenase (Psed_0244) and tyrosine with gentisate 1,2-dioxygenase protein (Psed_1337) and 4-hydroxyphenylpyruvate dioxygenase (Psed_1410). Genes annotated to be involved in the degradation of the following compounds were identified in the strain CB1190 genome: protocatechuate (Psed_1566 to Psed_1564), catechol (Psed_3093 to Psed_3095); vanillate (Psed_2465 to Psed_2466). A number

of genes suggesting that assimilation of benzoate may occur via benzoyl-CoA were discovered, such as those annotated as benzoate-CoA ligase (Psed_1491), benzoyl-CoA-dihydrodiol lyase (Psed_1490), and benzoyl-CoA oxygenase component B (Psed_1489).

A few ORFs annotated as dioxygenases are co-localized with BMM cluster genes involved in hydroxylation of aromatic compounds. A catechol 1,2-dioxygenase gene (Psed_0764) is found downstream of the phenol monooxygenase gene cluster (Psed_0770 to Psed_0766). Another ORF annotated as a catechol 1,2-dioxygenase (Psed_6065) is located immediately downstream of the HPA monooxygenase gene cluster (Psed_6067 to Psed_6066) and a few genes upstream of the aromatic monooxygenase 4 gene cluster (Psed_6062 to Psed_6058). The co-location of these BMMs and the catechol 1,2-dioxygenase genes is consistent with the fact that the metabolism of aromatic compounds (such as toluene, phenol, and HPA) typically involves monooxygenase reactions that produce dihydroxy aromatic compounds (such as catechol, benzoate, and dihydroxyphenylacetate). These dihydroxy aromatic compounds are substrates of ring-cleaving dioxygenase reactions, which produce intermediates that can enter the TCA cycle (27, 31).

Discussion

P. dioxanivorans strain CB1190 is the first bacterium with a finished genome in the genus *Pseudonocardia*. In general, members of the family *Pseudonocardiaceae* are characterized with a high GC-content (67-73%) and a large genome size (4.3-10.2 Mb). Among the *Pseudonocardiaceae* genomes, strain CB1190 (along with *Pseudonocardia* sp. P1) has the highest GC content (73%) but is medial in terms of its total genome size (7.1 Mb) and CDS count (6,797). The *Pseudonocardia* genomes, in general, share a large number of protein encoding genes with the genomes of *A. mediterranei* (3,035-3,219 CDS).

The majority of the sequenced *Pseudonocardiaceae* genomes are from organisms with industrial and medical importance, with the primary purpose to determine the proteins involved in the secondary metabolism of antibiotic production in these organisms. In strain CB1190, the only secondary metabolism-associated genes that we identified were genes encoding for antibiotic biosynthesis monooxygenases, beta-lactamases, polyketide cyclase/dehydrases, and proteins necessary for the production of precursors (*e.g.*, malonyl-CoA and methylmalonyl-CoA) used in secondary metabolism.

The presence of a large number of genes for antibiotic resistance in the CB1190 genome suggests that strain CB1190 evolved in a complex and competitive microbial environment with antibiotic producing organisms. Genes encoding a wide variety of drug efflux proteins, including multidrug transport proteins and those putatively targeting antibiotics such as acriflavin, actinorhodin, camphor, cephamycin, chloramphenicol, landomycin, methicillin, and rifamycin, were identified. In addition to transporting mechanisms for coping with antimicrobial metabolites, half of the putative dioxygenase genes spread across the genome are related to the bleomycin resistance protein, which could degrade antibiotics into less toxic products. The genome also harbors genes for transport and signal transduction mechanisms to cope with heavy metal toxicity. Therefore, in addition to dealing with a competitive microbial community, strain CB1190 appears to be well-suited to deal with a hostile environment, as would be expected of its isolation source: dioxane contaminated sludge. However, the nature and extent of heavy metal toxicity that strain CB1190 can withstand has not been studied. Characterizing the effect of heavy metals on contaminant degradation by strain CB1190 could determine the feasibility of bioremediation at dioxane sites contaminated with heavy metals.

Previous studies have implicated monooxygenases in catalyzing the first step of dioxane degradation in strain CB1190 (89, 91) (see Chapters 5 and 6 for experimental evidence). Monooxygenase-specific acetylene inhibition of THF and dioxane degradation by THF-grown CB1190 and *P. tetrahydrofuranoxydans* had demonstrated that a putative but biochemically uncharacterized THF monooxygenase is involved in the degradation of dioxane by *P. tetrahydrofuranoxydans* and simultaneously suggested that strain CB1190 possesses a homologous monooxygenase enzyme (91). The THF monooxygenase genes *thmADBC* located on a plasmid in *P. tetrahydrofuranoxydans* encode for a four subunit enzyme homologous to binuclear-iron-containing BMMs (154). Although a nearly identical *thm* gene cluster was found on plasmid pPSSED02, strain CB1190 contains seven additional BMMs that could potentially be involved in dioxane degradation. The aromatic monooxygenases 1-4 (Fig. 4.1) share a high degree of similarity with toluene monooxygenases. In other organisms toluene-2-monooxygenase (T2MO), toluene-*p*-monooxygenase (TpMO), and toluene-4-monooxygenase (T4MO) were found to degrade dioxane, but not toluene-*o*-xylene-monooxygenase (ToMO) (89). Interestingly, aromatic monooxygenases 2 and 4 are similar in sequence similarity and gene order (*tmoABCDE*) to both T4MO in *Pseudomonas mendocina* KR1, which can degrade dioxane, and ToMO in *Pseudomonas strutzeri* OX1, which cannot degrade dioxane. The gene clusters of aromatic monooxygenases 1 and 3 also contain T4MO and ToMO homologous genes but the arrangement of genes in these clusters is unique, *tmoABDEC*. In fact, the gene neighborhoods downstream of aromatic monooxygenases 1 and 3 are highly identical, suggesting the likelihood of a large duplication event in the genome of at least 5.9 kb. Toluene-grown *P. tetrahydrofuranoxydans* also degrades dioxane, but the type or sequence of the toluene-induced monooxygenase has not been determined (89). Although a propane-induced monooxygenase degrades dioxane in *Mycobacterium vaccae* JOB5 (19), suggesting that the propane monooxygenase identified in strain CB1190 could oxidize dioxane, we did not detect a propane monooxygenase in *M. vaccae* JOB5 using degenerate oligonucleotide PCR primers (Chapter 2). Therefore, a role for the strain CB1190 propane monooxygenase homologue in dioxane degradation is unclear. Sequence analysis and the arrangement of their genes indicate that the *prm* and *thm* clusters in strain CB1190 are closely related.

Conclusion

The genome of *Pseudonocardia dioxanivorans* strain CB1190 represents the first complete and annotated sequence of both a dioxane-metabolizing microorganism and a representative of the genus *Pseudonocardia*. This annotation reveals the potential for strain CB1190 to use a wide variety of organic substrates, as is typical for environmental actinobacteria. The genome provides a foundation for future studies to explore the genetic and biochemical basis of strain CB1190 physiology, such as dioxane and C2 carbon metabolism, as explored in Chapter 5 and 7, respectively.

5. Transcriptional analysis of *Pseudonocardia dioxanivorans* strain CB1190 dioxane degradation

This is a summary of work published in:

1. Grostern, A., C.M. Sales, W.-Q. Zhuang, O. Erbilgin, and L. Alvarez-Cohen. 2012. Glyoxylate metabolism is a key feature of the metabolic degradation of 1,4-dioxane by *Pseudonocardia dioxanivorans* strain CB1190. *Applied & Environmental Microbiology*, in press.
2. Sales, C. M. 2012. Functional genomics of bacterial degradation of the emerging contaminants 1,4-dioxane and N-nitrosodimethylamine (NDMA). Ph.D. thesis. University of California, Berkeley.
3. Additional results presented here are in publications in preparation.

Introduction

The cyclic ether dioxane is part of the growing class of emerging contaminants, whose wide distribution in the environment has only recently been recognized. Dioxane is highly soluble in water and thus is easily transported in groundwater systems from point source contamination (181). Although the long-term effects of dioxane exposure on human health are not well understood, animal studies indicate the compound is a potential carcinogen (146). Consequently, dioxane contamination has become the focus of remediation research, and microbial systems capable of dioxane degradation have been investigated (181). The degradation of dioxane under co-metabolic conditions (where other substrates provide carbon and energy for growth) appears to be somewhat common (19, 89, 101, 141, 148, 161), but an increasing number of microorganisms have also been identified that can use dioxane as a sole source of carbon and energy (13, 66, 89, 101, 112).

Initial investigations of biological dioxane transformation were performed as part of toxicological studies, which identified hydroxyethoxyacetic acid (HEAA) and dioxane-2-one as the major metabolites in humans (178) and rats (168), respectively. Recently, the pathways of dioxane biotransformation have been studied in a number of metabolizing and co-metabolizing bacteria and fungi in an attempt to understand the fate of environmentally-released dioxane. Similar to the findings of the toxicological studies, a common feature of dioxane transformation is the initial hydroxylation of the dioxane ring by monooxygenase enzyme systems (89, 141, 161). The hydroxylated metabolite then undergoes uncharacterized biotic or abiotic processing, leading to an opening of the ring structure. Researchers have used a number of techniques to detect post-hydroxylation metabolites in order to elucidate the dioxane catabolic pathway. In the dioxane-co-metabolizing fungus *Cordyceps sinensis*, deuterated dioxane-*d*₈ and GC-MS analysis were used to identify ethylene glycol, glycolic acid and oxalic acid as metabolites (101). In *Mycobacterium* sp. PH-06, which grows using dioxane, deuterated dioxane and GC-MS analysis detected only two metabolites, 1,4-dioxane-2-ol and ethylene glycol (66). Meanwhile, 2-hydroxyethoxyacetic acid (2HEAA) was the sole metabolite detected with ¹⁴C-labeled dioxane in the co-metabolizing bacterium *Pseudonocardia* strain ENV478 (161).

A more comprehensive analysis of dioxane metabolites was recently performed with the dioxane-metabolizing *Pseudonocardia dioxanivorans* strain CB1190 (designated strain CB1190) (91). By analyzing metabolites with Fourier transform ion cyclotron resonance-mass spectrometry, MS-MS, and by tracking distribution of ¹⁴C-labeled dioxane metabolites, the previously identified metabolites were confirmed, and 2-hydroxyethoxyacetaldehyde, 1,2-

dihydroxyethoxyacetic acid, 2-hydroxyethoxy-2-hydroxyacetic acid, glycoaldehyde, glyoxylic acid and CO₂ were also identified as dioxane metabolites (91). Based on these findings, a dioxane biodegradation pathway was proposed (Fig. 5.1, black arrows), whereby all metabolites feed through oxalic acid into central metabolism (91).

These studies have improved our understanding of the fate of dioxane during microbial biodegradation, but there is a knowledge gap in terms of the biochemical and genetic basis for this pathway. While monooxygenase enzymes have been implicated in the initial activation of the dioxane ring, the further steps in dioxane metabolism have not been explored. In particular, the mechanism for incorporation of dioxane metabolites into biomass in microorganisms that can grow with dioxane as the sole carbon and energy source is unknown. In this work we used the recently-determined genome of *P. dioxanivorans* strain CB1190 (132) to investigate how dioxane is assimilated into biomass to support growth. We present evidence highlighting the unique role of glyoxylate in bacterial dioxane metabolism.

Materials and Methods

Culture growth

Strain CB1190 was cultivated in ammonium mineral salts (AMS) medium (112). Replicate were amended with the appropriate carbon source. These were then inoculated with 1/500 dilution dioxane-grown strain CB1190 culture and were incubated at 30°C with shaking. Sterile (un-inoculated) controls were prepared in parallel. Bottles were amended with dioxane, glycolate, or pyruvate. Glycolate was chosen since, of the known dioxane degradation intermediates, it best supported strain CB1190 growth.

Cell harvesting and RNA extraction for transcription studies

Cells were harvested when approximately half of the added substrate had been consumed. Cells from replicate bottles were collected by filtration onto triplicate filters, were scraped from the filters with a sterile scalpel, and were transferred to 2 mL screw-top microcentrifuge tubes containing zirconia-silica beads (Biospec Products, Bartlesville, OK). Harvested cells were stored at -80°C until RNA extraction.

Nucleic acids were extracted using a previously published phenol method (60). To obtain RNA, resuspended nucleic acids were initially separated with the Allprep kit (Qiagen), and then the RNA was purified using the RNeasy kit (Qiagen). Following elution with RNase-free water, any contaminating DNA were removed by two successive DNase I treatments using the DNA-free kit (Ambion, Austin, TX) according to the manufacturer's instructions. Pure RNA was obtained with a final cleanup with the RNeasy kit.

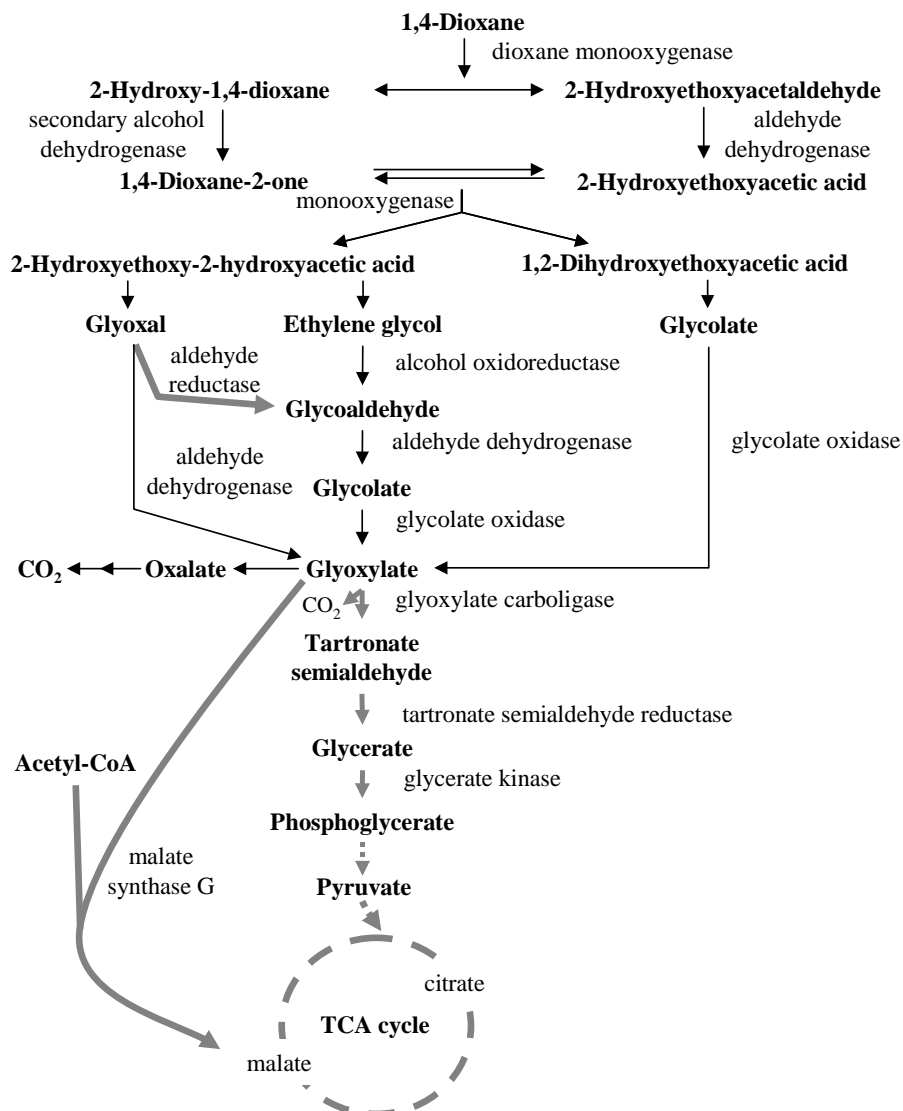


Figure 5.1. Proposed pathways and enzymes involved in dioxane metabolism in strain CB1190.

Black arrows indicate the transformations previously proposed by Mahendra *et al.* (91). Grey arrows indicate transformations supported by *in silico* or experimental results in the current work. Dashed arrows indicate multi-step transformations.

Microarray design and analyses

A draft version of the strain CB1190 genome was obtained from the Joint Genome Institute and consisted of 6896 protein-encoding sequences in 226 contigs. These were submitted to Affymetrix (Santa Clara, CA) for custom chip design. The final design consisted of a set of perfect match-only 25-mer probes, with each sequence targeted by 11-13 probes, and also included a standard set of Affymetrix controls for prokaryotic gene expression microarrays. Subsequent to microarray production, a finished genome sequence for strain CB1190 was made available (132). The microarray probes were re-annotated to reflect the finished genome sequence. Re-annotated microarray probe sets targeted 6391 (approximately 94%) of the coding sequences in the finished genome.

Five µg of total RNA was used as starting material for each microarray analysis. Triplicate microarrays were processed for each substrate treatment. cDNA was synthesized, fragmented, labeled, and hybridized to arrays according to the protocols outlined in section 3 of the Affymetrix GeneChip Expression Analysis technical manual, with the following changes: cDNA fragmentation was achieved by the addition of 0.07 U DNase I/µg cDNA, and the hybridization temperature was 52°C. Hybridized arrays were stained and washed according to Affymetrix protocol “Modified FlexMidi_euk2v3 for *P. aeruginosa* Array” and were scanned using an Affymetrix GeneChip Scanner 3000.

Microarray data was first processed with the RMAExpress software program (15) to compute gene expression summary values. Background adjustment and quantile normalization were performed across all microarrays. The average normalized expression ratio (dioxane or glycolate treatment/ pyruvate control) was calculated for each gene. Ratios were considered significant if they were >2.0 or <0.5 and Student's t-test indicated a *p*-value (two-tailed) of <0.05.

Quantitative RT-PCR

To verify microarray data, the relative transcription of 17 genes in the dioxane and glycolate treatments were compared with the pyruvate treatment. cDNA was synthesized from 500 ng of RNA using the TaqMan reverse transcription kit (Applied Biosystems, Foster City, CA) with random hexamers in 15 µL reactions according the manufacturer's protocol. cDNA was diluted 100-fold and 2 µL were used in each subsequent qPCR reaction. All qPCR reactions were performed in triplicate with an Applied Biosystems StepOne Plus real-time PCR system. For most genes, SYBR Green chemistry was used, and each 20 µL reaction consisted of 1X Fast SYBR Green master mix, 2 µL of diluted cDNA and each primer at 0.5 mM. The PCR cycling conditions were: 95°C for 20 s, then 40 cycles of (95°C for 3 s, annealing for 30 s), followed by melting curve analysis. The genes *tpi*, *thiC*, and *rpoD* were determined by the geNorm method (163) to be stably transcribed across all treatments, and were thus used as internal references. TaqMan chemistry was used for these genes, and each qPCR reaction consisted of 1X Fast Universal Mix (Applied Biosystems), 2 µL of diluted cDNA, each primer at 0.5 mM, and the probe at 145 nM. The cycling conditions were: 95°C for 20 s, then 40 cycles of (95°C for 1 s, annealing for 20 s). The primer sequences and dye and quencher chemistry for each probe are listed in Table 5.1. The efficiency of each qPCR assay was determined with a serial dilution of cDNA derived from a dioxane treatment replicate. Gene expression was normalized using the method of Vandesompele et al. (163).

Table 5.1. SYBR Green and TaqMan primers used in qRT-PCR analyses.^{a,b}

Target Gene ID	Gene product	Sense primer sequence	Antisense primer sequence	Probe sequence
Psed_0629	Methane monooxygenase	GAACCCGGAGCTGCACAA	TTCAGGTTCTGCTGGATCGT	-
Psed_0768	Phenol 2-monooxygenase	GCTCTTCGTGCCCTTCATGA	GAGAACCCGAACGACATCGT	-
Psed_0815	Methane/phenol/toluene hydroxylase	GGTGAAGGACCGTGACAAC	GGCCAGGAAGCGATCGAT	-
Psed_3051	RNA polymerase sigma factor RpoD	TGATCCAGGAGGGCAACCT	TAGCCCTTGGTGTAGTCGAACTT	(TET)CTGATCCGTGCGGTC(TAMRA)
Psed_3417	Triosephosphate isomerase	CCGGCAACTGGAAGATGAAC	GGCGAGCTTCTGCAGGAA	(VIC)CACCTCGAGGCCATC(TAMRA)
Psed_3888	Hydroxyruvate isomerase	CTTCTGTTCGACCTGTCCA	TGCACGTGGGCGATCAC	-
Psed_3889	Tartronate semialdehyde reductase	ACATGCTGAACCGGGACTTC	CTTCATGTCTTGTGGTGCAA	-
Psed_3890	Glyoxylate carboligase	GCGCACACGCACTACGACC	CGGCCGAAGACCCGGTTCATC	-
Psed_3891	Glycerate kinase	GATCGAGCAGGCCTACGC	TGCGAGCTCCTCCAGCAG	-
Psed_4782	Malate synthase	CGAGCAGACGGTCCGGGAGA	GTCGCTCATCGGCCGGTAGC	-
Psed_4783	Glycolate oxidase	GGCCGCATCTCCCCGACTA	CTCGCCCATGTCCGCGATCC	-
Psed_4788	Glycolate oxidase	CCGACCTCGCGACGATGACG	GAGGGGTGGGCAGCAGCTTG	-
Psed_6168	Thiamine biosynthesis protein C	TGAAGGTCAACGCCAACATC	TGGCCCACACCAGTTTCTC	(FAM)CGTCGATCGAGGACG(TAMRA)
Psed_6782	Alcohol dehydrogenase	GCAGGTCGAGTCAAGGACTTTC	GCCGTGCGGCACCAT	-
Psed_6783	Fatty-acid-coA ligase	CTCTAGTGGACGAGGAGGAAGT	TCGCTTAGGCACCTTGTAACC	-
Psed_6970	FAD/FMN-dependent dehydrogenase	GAGTTGACTTCGATCTGCTTACGA	GGCCGCATCAACCATGAA	-
Psed_6971	Alcohol dehydrogenase	TTGCTGGGATGGTGAAGGA	CCATGAGGCACCATTGCTT	-
Psed_6975	Aldehyde dehydrogenase	TGACATCGAGCAGGCTGTTG	CAGGTCTGCCCGTTGTTGT	-
Psed_6976	Monooxygenase oxidoreductase domain	GGCTGGGACCGGACGTA	TTGTTCCGCAGCTTGTATCG	(TET)TACCCCCATGAGCGCGGGAA(TAMRA)
Psed_6981	Aldehyde dehydrogenase	GGAGCGCCGGTTATTATCAA	TGACCCAAGAACAGCGATGA	-

^a All primers sets were annealed at 60°C, except for Psed_4783, which was annealed at 65°C. ^b Except for where probe sequences are listed, all PCR reactions were performed with SYBR-Green chemistry

Cloning and expression of strain CB1190 genes

Strain CB1190 genes Psed_3889 and Psed_3890 were PCR amplified and ligated into linearized *Rhodococcus* expression plasmid pTip-QC2 (102) using the hetero-stagger cloning method (85). *Rhodococcus jostii* strain RHA1 was transformed with pTip-Psed_3889, pTip-Psed_3890, or pTip-QC2, and then expression of the inserted gene was induced by the addition of thiostrepton to the culture medium. Cell-free extracts from strain RHA1/pTip cells were prepared as described below for strain CB1190 cells.

Glyoxylate carboligase assay

Strain CB1190 cells grown with either pyruvate or dioxane were harvested by filtration and scraped cells were stored at -80°C. Cell-free extracts were prepared by resuspending the cells in sodium phosphate buffer (10 mM, pH 7.0), adding zirconia-silica beads, and bead beating. Cell debris was collected by centrifugation, and the supernatant (cell-free extract) was aliquoted and stored at -80°C until use. Protein was determined by the method of Bradford (17).

Glyoxylate carboligase activity was tested using the linked glyoxylate carboligase/tartronate semialdehyde reductase assay described by Cusa et al. (28). Briefly, 2 mL reactions, containing 100 mM sodium phosphate (pH 8.0), 5 mM MgCl₂, 0.28 mM NADH, 0.5 mM thiamine pyrophosphate, and 200 µg of protein from cell-free extracts, were prepared in triplicate in screw-top glass vials with Mininert caps. The headspace was with N₂, glyoxylate was added to 10 mM, and vials were incubated with shaking at 30°C for 16 h. Reactions were stopped and CO₂ liberated by the addition of 10 N sulfuric acid and shaking for 15 min. CO₂ was analyzed by gas chromatography-pulsed discharge detection (GC-PDD) and glyoxylate was analyzed by HPLC.

For assays with cell-free extracts from strain RHA1 carrying pTip plasmid constructs, the assay vials were prepared as above for the following treatments: 1) pTip-Psed_3889 (8 µg of protein); 2) pTip-Psed_3890 (7 µg of protein); 3) pTip-Psed_3889 + pTip-Psed_3890 (8 µg and 7 µg of protein of each, respectively); 4) pTip-QC2 (10 µg of protein); 5) Buffer control. CO₂ production was determined by GC-PDD, and glyoxylate consumption and glycerate production were determined by HPLC.

Growth with 1,4-[U-¹³C]dioxane

A series of strain CB1190 cultures was prepared with 1,4-[U-¹³C]dioxane (99% pure; Sigma-Aldrich). Serum bottles with black butyl rubber stoppers and containing AMS medium were amended with ¹³C-dioxane, or ¹³C-dioxane plus 3% CO₂. These bottles were inoculated with 1/300 dilution ¹³C-dioxane-grown strain CB1190 culture and were incubated at 30°C with shaking. Following complete removal of the dioxane, cells from each treatment were transferred to fresh triplicate bottles with the same amendments and were incubated with shaking until dioxane removal was complete. Cells were then harvested by filtration, scraped from the filter, and stored at -80°C until isotopomer amino acid analysis.

Analytical methods

Pyruvate, glycolate and glyoxylate consumption and glycerate production were monitored with a Waters (Milford, MA) model 2695 HPLC equipped with a model 2996 photodiode array detector and an Amidex HPX-87H column (Bio-Rad, Hercules, CA). Dioxane consumption was monitored by direct injection of samples onto a Varian 3400 GC equipped with a FID detector

and a 1% AT-1000 on carbograph1 packed column (Grace, Columbia, MD). CO₂ production in cell-free extracts was determined by injecting headspace samples onto a HP 6890 GC equipped with a PDD detector, a 1.5 m Hayesep-DB 100/120 pre-column and a 2 m Hayesep-DB 120/140 column.

The preparation and isotopomeric analysis of proteinogenic amino acids were performed as previously described. Two types of charged fragments were clearly detected by GC-MS for the derivatized amino acids: the [M-57]⁺ which contains the entire amino acid; and the [M-159]⁺, which contains the amino acid without the first carbon (α carboxyl group). The [M-57]⁺ peaks in leucine, isoleucine and proline overlap with other peaks, so the [M-159]⁺ group was used to obtain the isotopomer labeling information of those amino acids. The final isotopomer labeling fractions were indicated as: M0 (unlabeled fraction), M1 (single labeled carbon fraction), M2 (fraction with two labeled carbons), M3 (fraction with three labeled carbons), and so forth.

Results and Discussion

In silico analysis of dioxane degradation pathway genes

The pathway for dioxane degradation that was proposed by Mahendra *et al.* (91) (Fig. 5.1, black arrows) was the starting point for this study. We searched the literature for descriptions of enzymatic activities catalyzing the proposed dioxane metabolic transformations, and used the amino acid sequence of characterized enzymes to identify homologues in the strain CB1190 genome sequence using blastp (4).

Dioxane degradation is initiated by the activity of a monooxygenase (89, 161). The strain CB1190 genome carries eight bacterial multi-component monooxygenase gene clusters, seven of which are encoded on the chromosome and one that is on plasmid pPSED02 (132).

The transformation of the initial hydroxylated dioxane metabolite is proposed to involve spontaneous ether cleavage and/or aldehyde/carboxylic acid and/or secondary alcohol/ketone conversions (91). Strain CB1190 has 13 genes encoding proteins with at least 30% amino acid identity with characterized bacterial aldehyde dehydrogenases (*E. coli* AldA [accession NP_415933], *E. coli* AldB [accession AAC76612], *Rhodococcus erythropolis* AldH [accession AAZ14956]), which catalyze the NAD(P)⁺-dependent oxidation of aldehydes to carboxylic acids. Eleven of these genes are located on the chromosome, while the remaining two are on plasmid pPSED02. Three genes encoding proteins with 29-36% amino acid identity to a secondary alcohol dehydrogenase (accession CAD36475) of the actinomycete *Rhodococcus ruber* are on the chromosome; this enzyme catalyzes the NAD⁺-dependent transformation of secondary alcohols to ketones (89, 161).

The second proposed hydroxylation reaction generating single ether-containing 2-hydroxyethoxy-2-hydroxyacetic acid or 1,2-dihydroxyethoxyacetic acid has also been postulated to be catalyzed by a monooxygenase (91) (see Fig. 5.1).

The subsequent cleavage of the second ether bond leads to the production of two-carbon intermediates, of which glyoxal, ethylene glycol, glycoaldehyde, glycolate, glyoxylate and oxalate have been identified (91). The strain CB1190 genome has three genes encoding proteins with 27-34% amino acid identity to *E. coli* 1,2-propanediol oxidoreductase (accession AP_003365), which transforms ethylene glycol to glycoaldehyde (26). Glycoaldehyde is transformed to glycolate by aldehyde dehydrogenases (see above), while glycolate is converted to glyoxylate by glycolate oxidase (70). In *E. coli*, glycolate oxidase is encoded by *glcDEF* (116); strain CB1190 has two adjacent homologues, *glcD1E1F1* and *glcD2E2F2* (Fig. 5.2).

These homologues share 72%, 47%, and 64% amino acid identity for GlcD, GlcE and GlcF, respectively.

Mahendra et al. (91) proposed that glyoxal is transformed to glyoxylate, and although this enzymatic activity has not been previously described in the literature, the numerous aldehyde dehydrogenases encoded in the strain CB1190 genome could potentially catalyze this reaction. Alternatively, genes encoding aldehyde reductase homologues (*E. coli* YqhD [accession Q46856], *Bacillus subtilis* YvgN [accession O32210]) that catalyze the transformation of glyoxal to glycoaldehyde were identified. Furthermore, in *E. coli* the glyoxalase I/II system, encoded by the genes *gloA* and *gloB*, uses S-glutathione to transform glyoxal to glycolate (155); however, no homologous genes were identified in the strain CB1190 genome.

Since glyoxylate has been identified as a metabolite of dioxane (91, 101), we examined the strain CB1190 genome for genes involved in known glyoxylate assimilation pathways. Strain CB1190 has a cluster of genes (Fig. 5.2) homologous to those encoding the glyoxylate carboligase pathway that converts glyoxylate through tartronate semialdehyde and glycerate to phosphoglycerate, in order to provide both carbon and energy for growth in *E. coli* (73). Strain CB1190 also has a gene (Psed_4782) that encodes for malate synthase G, which in *E. coli* (and in strain CB1190 – see Fig. 5.2) is linked to the locus encoding glycolate oxidase and which assimilates glyoxylate directly into the TCA cycle via malate (71, 99). Malate synthase G is non-essential for the growth of *E. coli* on glyoxylate, and is therefore considered to play an anaplerotic role (110).

While the transformation of glyoxylate to oxalate has been reported in *Pseudomonas fluorescens* (140), the gene sequence for the enzyme catalyzing this reaction has not yet been determined in bacteria. Finally, aerobic bacterial growth on oxalate was shown to be dependent on the oxalate decarboxylase OxdC, which converts oxalate to formate (150). However, the strain CB1190 genome lacks a homologue for this enzyme.

Gene expression during growth on dioxane, glycolate and pyruvate

We used gene expression microarrays targeting the strain CB1190 genome to determine the genes that are differentially regulated during growth on dioxane or glycolate relative to growth on pyruvate. The expression of 383 genes differed significantly; 97 genes were up-regulated with dioxane relative to the pyruvate control, whereas 286 genes were down-regulated. When strain CB1190 was grown with glycolate, the expression of 506 genes differed significantly relative to pyruvate-grown cells, with 203 genes up-regulated and 303 down-regulated. A comparison of genes differentially regulated with either dioxane or glycolate relative to pyruvate identified 36 genes that were up-regulated with both substrates, whereas 121 genes were down-regulated with both substrates. Table 5.2 provides a list of genes that were up-regulated with dioxane alone or with both dioxane and glycolate.

The replicon location of the dioxane- and glycolate-induced genes was examined. For dioxane-grown cells, 27 and 17 genes were induced from plasmids pPSED01 and pPSED02, respectively, with the remaining induced genes on the chromosome. For glycolate-grown cells, 15 induced genes were from plasmid pPSED01, while the rest were on the chromosome; no genes were induced from plasmid pPSED02.

Sixty-one genes were up-regulated with dioxane but not with glycolate, relative to pyruvate (Table 5.2). Of these genes, 21 were from plasmid pPSED01, while 17 genes were from plasmid pPSED02. Of the dioxane-induced plasmid pPSED01 genes, 14 of these encode hypothetical proteins.

Table 5.2. Strain CB1190 genes on microarray up-regulated during growth with dioxane^a

Gene ID	Gene name	Gene product	Replicon ^b	Expression ratio ^c	
				Dioxane/ Pyruvate	Glycolate/ Pyruvate
Psed_0038		Regulatory protein ArsR	c	2.0	0.37
Psed_0350		Bile acid:sodium symporter	c	2.9	13
Psed_1076		Short-chain dehydrogenase/reductase SDR	c	2.6	1.7
Psed_1302		Protein of unknown function DUF156	c	3.2	11
Psed_1303		Heavy metal transport/detoxification protein	c	2.7	8.0
Psed_1304		Heavy metal translocating P-type ATPase	c	2.5	9.7
Psed_1584		6-Phosphofructokinase	c	3.3	3.9
Psed_1594		Protein of unknown function UPF0016	c	2.3	1.2
Psed_1658		Trimethylamine-N-oxide reductase (cytochrome c)	c	5.3	0.63
Psed_2030		Linalool 8-monooxygenase	c	2.2	2.4
Psed_2371		ABC-type transporter, periplasmic subunit	c	2.0	3.1
Psed_3041		NADH dehydrogenase (ubiquinone) 24 kDa subunit	c	2.0	1.1
Psed_3522		Response regulator receiver	c	3.5	3.9
Psed_3564		Aromatic-amino-acid transaminase	c	3.0	1.2
Psed_3888	<i>hyi</i>	Hydroxy pyruvate isomerase	c	106	112
Psed_3889	<i>glxR</i>	Tartronate semialdehyde reductase	c	114	114
Psed_3890	<i>gcl</i>	Glyoxylate carboligase	c	98	108
Psed_3891	<i>glxK</i>	Glycerate kinase	c	49	56
Psed_3934		Major facilitator superfamily MFS_1	c	2.4	1.6
Psed_3935		GntR domain protein	c	2.5	1.1
Psed_4031		Iron-sulfur cluster binding protein	c	2.1	2.8
Psed_4146		Potassium-transporting ATPase B chain	c	2.6	1.2
Psed_4147		Potassium-transporting ATPase A chain	c	2.6	1.2
Psed_4148		K+-transporting ATPase, F subunit	c	3.0	1.3
Psed_4512		ABC-type transporter, integral membrane subunit	c	2.2	2.7
Psed_4513		Cobalamin (vitamin B12) biosynthesis CbiX protein	c	2.1	2.4
Psed_4577		Cold-shock protein DNA-binding	c	2.2	3.7
Psed_4755		Ammonium transporter	c	2.2	9.3
Psed_4782	<i>glcB</i>	Malate synthase	c	10	11
Psed_4788	<i>glcD2</i>	D-lactate dehydrogenase (cytochrome)	c	43	44
Psed_4789	<i>glcE2</i>	FAD linked oxidase domain protein	c	24	24
Psed_4790	<i>glcF2</i>	Protein of unknown function DUF224 cysteine-rich region domain protein	c	27	26
Psed_5025		4-Hydroxyacetophenone monooxygenase	c	5.8	3.6
Psed_5026		Regulatory protein TetR	c	2.4	2.8
Psed_5135		Malic protein NAD-binding	c	2.6	1.4
Psed_5406		NAD(P)(+) transhydrogenase (AB-specific)	c	2.1	2.2
Psed_5524		Hydrolase	c	2.1	1.3
Psed_5736		Transglycosylase-like domain protein	c	2.8	8.4
Psed_5938		Glycoside hydrolase family 13 domain-containing protein	c	2.3	1.5
Psed_6175		Regulatory protein TetR	c	3.5	4.0
Psed_6259		FMN-dependent oxidoreductase, nitrilotriacetate monooxygenase family	c	2.5	1.2
Psed_6261		ABC-type transporter, integral membrane subunit	c	2.3	1.1
Psed_6262		ABC-type transporter, integral membrane	c	2.2	0.97

	subunit			
Psed_6263	Extracellular ligand-binding receptor	c	2.2	0.92
Psed_6600	Acyl-CoA dehydrogenase domain-containing protein	c	4.2	0.95
Psed_6728	Amidase	p1	2.6	0.99
Psed_6730	Luciferase-like, subgroup	p1	2.8	0.72
Psed_6732	MaoC domain protein dehydratase	p1	2.2	1.0
Psed_6735	Flavoprotein WrbA	p1	2.1	1.1
Psed_6742	NLP/P60 protein	p1	2.7	1.1
Psed_6745	ATP-binding protein	p1	3.7	0.72
Psed_6751	Transglycosylase-like domain protein	p1	5.0	0.60
Psed_6779	Integrase catalytic region	p1	6.9	5.4
Psed_6782	Hydroxy acid-oxoacid transhydrogenase	p1	4.6	4.3
Psed_6784	Alkylglycerone-phosphate synthase	p1	3.5	3.5
Psed_6787	Formyl-CoA transferase	p1	13	11.3
Psed_6791	IstB domain protein ATP-binding protein	p1	3.5	2.4
Psed_6799	Transposase IS 4 family protein	p1	2.0	1.7
Psed_6970	D-lactate dehydrogenase (cytochrome)	p2	6.3	0.67
Psed_6971	Hydroxyacid-oxoacid transhydrogenase	p2	5.1	0.69
Psed_6972	GntR domain protein	p2	3.4	0.62
Psed_6974	Ethyl tert-butyl ether degradation EthD	p2	3.2	0.71
Psed_6975	Betaine-aldehyde dehydrogenase	p2	2.6	0.74
Psed_6977	<i>thmD</i> Ferredoxin--NAD(+) reductase	p2	4.2	0.72
Psed_6978	<i>thmB</i> Methane/phenol/toluene hydroxylase	p2	2.7	0.80
Psed_6979	<i>thmC</i> Monooxygenase component MmoB/DmpM	p2	3.7	0.76
Psed_6981	Aldehyde Dehydrogenase	p2	9.5	0.83
Psed_6982	Mn2+/Fe2+ transporter, NRAMP family	p2	14	1.0
Psed_7002	Transcription factor WhiB	p2	2.7	0.67

^a Only annotated genes with predicted functions are shown. The following genes (Psed_XXXX) were automatically annotated as “hypothetical protein”: (chromosome) **2752, 3653, 3669, 4149, 4306, 4424, 5371**; (plasmid pPSED01) **6729, 6743, 6744, 6746, 6747, 6748, 6749, 6750, 6752, 6753, 6754, 6755, 6757, 6758, 6780**; (plasmid pPSED02) **6913, 6973, 6980, 7003, 7007, 7008**

^b c = chromosome; p1 = plasmid pPSED01; p2 = plasmid pPSED02

^c Genes up-regulated with dioxane but not with glycolate are indicated in **bold**

Upregulation of a multi-component monooxygenase gene cluster during dioxane metabolism

While the strain CB1190 genome encodes eight monooxygenase gene clusters, microarray analysis of dioxane-induced transcription revealed that only the plasmid pPSED02-encoded monooxygenase cluster was up-regulated when compared with growth with pyruvate (Table 5.2). Upregulation of this cluster was verified with qRT-PCR using primers targeting Psed_6976, encoding a homologue of THF monooxygenase α -subunit (Table 5.3). Three chromosomally-encoded monooxygenase gene clusters were missing from the microarray, so their potential involvement in dioxane metabolism was tested using qRT-PCR with primers targeting the genes encoding the α -subunits. As Table 5.3 shows, no significant differential transcription was observed for these three monooxygenase genes (Psed_0629, Psed_0768, and Psed_0815) in cDNA from dioxane-, glycolate- or pyruvate-grown cells.

The dioxane-induced, plasmid pPSED02-endoded monooxygenase gene cluster is homologous to the *thm* gene cluster involved in THF utilization in *Pseudonocardia tetrahydrofuranoxydans* strain K1 (154). This gene cluster has also been identified in DNA extracted from *Pseudonocardia* strain ENV478 (93), and in both strains K1 and ENV478, growth with THF induces transcription of genes in this cluster (93, 154). Although strains K1 and ENV478 degrade dioxane co-metabolically when induced with THF (89, 161), the published results did not directly link the THF-induced monooxygenase gene cluster to biochemical dioxane transformation activity. However, the up-regulation of the homologous *thm* monooxygenase gene cluster in strain CB1190 during dioxane metabolism reported here supports the involvement of the *thm* monooxygenase gene clusters in dioxane degradation in strain K1 and ENV478.

Table 5.3. Normalized relative expression levels for selected strain CB1190 genes analyzed by qRT-PCR.

	Gene ID	Gene name	Gene product	Replicon ^a	Expression ratio ^b	
					Dioxane/ Pyruvate	Glycolate/ Pyruvate
Monooxygenase genes lacking probes on microarray	Psed_6976	<i>thmA</i>	THF monooxygenase α-subunit	p2	15 \pm 0.2	1.0 \pm 0.5
	Psed_0629	<i>prmA</i>	Methane monooxygenase	c	0.4 \pm 0.1	0.8 \pm 0.5
	Psed_0768		Phenol 2-monooxygenase	c	2 \pm 2	0.9 \pm 0.4
	Psed_0815		Methane/phenol/toluene hydroxylase	c	1.1 \pm 0.5	0.3 \pm 0.4
Dioxane metabolism intermediates-related genes	Psed_6782		Alcohol dehydrogenase	p1	7.3 \pm 0.8	3.3 \pm 0.4
	Psed_6783		Fatty-acid-CoA ligase	p1	24 \pm 2	7 \pm 5
	Psed_6970		FAD/FMN-dependent dehydrogenase	p2	6.1 \pm 0.8	1.5 \pm 0.9
	Psed_6971		Alcohol dehydrogenase	p2	8.7 \pm 0.8	1.3 \pm 0.4
	Psed_6975		Aldehyde dehydrogenase	p2	23 \pm 2	0.9 \pm 0.3
	Psed_6981		Aldehyde dehydrogenase	p2	11 \pm 1	1.5 \pm 0.9
Glycolate and glyoxylate metabolism-related genes	Psed_4783	<i>glcD</i> 1	Glycolate oxidase	c	0.3 \pm 0.1	0.7 \pm 0.5
	Psed_4788	<i>glcD</i> 2	Glycolate oxidase	c	3.6 \pm 0.3	2.8 \pm 0.5
	Psed_4782	<i>glcB</i>	Malate synthase G	c	5.4 \pm 0.2	2 \pm 1
	Psed_3888	<i>hyi</i>	Hydroxypyruvate isomerase	c	5 \pm 1	3.1 \pm 0.3
	Psed_3889	<i>glxR</i>	Tartronate semialdehyde reductase	c	13 \pm 1	6 \pm 2
	Psed_3890	<i>gcl</i>	Glyoxylate carboligase	c	18 \pm 2	10 \pm 4
	Psed_3891	<i>glxK</i>	Glycerate kinase	c	1.9 \pm 0.6	1.9 \pm 0.7

^a c = chromosome; p1 = plasmid pPSED01; p2 = plasmid pPSED02

^b Genes up-regulated with dioxane but not with glycolate are indicated in **bold**

Genes potentially contributing to transformation of C₄ dioxane metabolites

After the initial monooxygenase-catalyzed hydroxylation, the C₄ dioxane metabolites undergo a series of transformations that ultimately yield two C₂ compounds (Fig. 5.1). The set of genes on plasmid pPSED02 that are induced by dioxane but not by glycolate could potentially be involved in these transformations (Table 5.2). The operon containing the *thm* monooxygenase-encoding gene cluster also includes genes encoding two aldehyde dehydrogenases (Psed_6975 and Psed_6981), and upstream of the monooxygenase cluster and in the reverse orientation is an upregulated gene cluster encoding an alcohol dehydrogenase and a FAD/FMN-dependent dehydrogenase, as well as a transcriptional regulator (Psed_6970-6972). Dioxane-dependent upregulation of these genes was verified by qRT-PCR (Table 5.3). The aldehyde-carboxylic acid conversion of C₄ dioxane metabolites could be catalyzed by these dioxane-induced aldehyde dehydrogenases. In contrast, a candidate gene encoding an enzyme for the secondary alcohol-ketone conversion of C₄ metabolites (*i.e.*, 2-hydroxy-1,4-dioxane to 1,4-dioxane-2-one) is not immediately clear, since none of the secondary alcohol dehydrogenase genes identified by *in silico* analysis were specifically up-regulated with dioxane. The alcohol dehydrogenase-encoding Psed_6971 or constitutively expressed secondary alcohol dehydrogenases (not identified in the differential-transcription analyses reported here) could be involved in this reaction. We are currently cloning the up-regulated dehydrogenases and reductases to functionally test their role in the transformation of C₄ dioxane metabolites.

The dioxane metabolic pathway also requires a second hydroxylation and two ether cleavage steps. Ether cleavage may occur spontaneously following hydroxylation, or it may occur enzymatically (167). A heterogeneous group of ether cleaving enzymes have been identified (reviewed in White et al. (167)), and an ether-bond cleaving diglycolic acid dehydrogenase has been purified from *P. tetrahydrofuranoxydans* strain K1 (172). However, genomic and transcriptomic analysis of dioxane metabolism did not identify such enzymes in strain CB1190.

Genes potentially contributing to transformation of C₂ dioxane metabolites to glyoxylate

The plasmid pPSED02-encoded putative aldehyde and alcohol dehydrogenases that were up-regulated during growth on dioxane but not on glycolate could also contribute to the transformation of the dioxane metabolites glyoxal, glycoaldehyde and ethylene glycol. In *E. coli* the aldehyde reductase YqhD catalyzes the NADPH-dependent reduction of glyoxal to glycoaldehyde (81), while in *Bacillus subtilis* the non-homologous YvgN catalyzes the same reaction (130). The up-regulated aldehyde dehydrogenase gene products share only low amino acid sequence identity with these proteins, but Psed_6975 has 35% identity to *E. coli* AldA, which catalyzes the NAD⁺-dependent oxidation of glycoaldehyde to glycolate (53). Boronat *et al.* (16) identified an *E. coli* mutant that employed the activity of propanediol oxidoreductase (*fucO* (26)) to transform ethylene glycol to glycoaldehyde, which was further transformed to glycolate. The putative alcohol dehydrogenase Psed_6971 that was up-regulated with dioxane shares 42% amino acid similarity to *E. coli* 1,2-propanediol oxidoreductase, so the gene product may catalyze the conversion of ethylene glycol to glycoaldehyde during dioxane metabolism. It should be noted that product of the glycolate- and dioxane-induced alcohol dehydrogenase gene Psed_6782 has similar amino acid identity to *E. coli* propanediol oxidoreductase, although its role in glycolate metabolism is not clear.

The set of genes up-regulated with both dioxane and glycolate relative to pyruvate included the putative glycolate oxidase-encoding chromosomal gene cluster *gldD2E2F2*

(Psed_4788-4790) (Table 5.2). In contrast, the adjacent homologous *glcDIE1F1* gene cluster was not up-regulated with either dioxane or glycolate. This lack of up-regulation was confirmed with qRT-PCR (Table 5.3). Therefore, the enzyme encoded by *glcD2E2F2* likely catalyzes the transformation of glycolate to glyoxylate (116). In *E. coli*, genes for malate synthase G (*glcB*) and *glcDEF* are proximal, oriented in the same direction, and are co-transcribed during growth with glycolate (117). While *glcB* (Psed_4782) and *glcD2E2F2* are in close proximity in strain CB1190, they have opposing orientation and are separated by the homologous but non-identical *glcDIE1F1* (Psed_4783-4785) cluster that was not up-regulated during growth with either dioxane or glycolate (Fig. 5.2). Between *glcD2E2F2* and *glcDIE1F1* is a gene (Psed_4787) encoding a putative GntR family transcriptional regulator, which may be involved in regulation of this glycolate oxidase homologue.

Glyoxylate metabolism during dioxane degradation by strain CB1190

Genes encoding two divergent routes for glyoxylate assimilation were up-regulated with both dioxane and glycolate, as determined by microarray analysis. The first route was through the *glcB*-encoded malate synthase G (Psed_4782) (Table 5.2), although qRT-PCR did not confirm this up-regulation (Table 5.3). The up-regulated gene cluster encoding glyoxylate carboligase, tartronate semialdehyde reductase, hydroxypyruvate isomerase and glycerate kinase (Psed_3888-3891; Fig. 5.2) represents the second route for glyoxylate assimilation. The up-regulation of three genes from this cluster was verified by qRT-PCR (Table 5.3).

These transcriptional results point to glyoxylate, rather than oxalate (Fig. 5.1), as the key intermediate in the assimilation of carbon into central metabolism during dioxane degradation by strain CB1190. In the glyoxylate carboligase pathway, two glyoxylate molecules are combined by glyoxylate carboligase, producing a C₃-compound that is eventually incorporated into glycolysis at phosphoglycerate (73) (Figs. 5.1 and 5.2). In the typical malate synthase G pathway, glyoxylate and acetyl-CoA are condensed to form malate, which is incorporated in the TCA cycle (71) (Figs. 5.1 and 5.2). As noted above, the glyoxylate carboligase pathway supports growth through carbon and energy conservation, whereas malate synthase G is involved in anaplerotic reactions but is not known to support growth in the absence of glyoxylate carboligase with glyoxylate as a sole substrate (110). Therefore, the growth of strain CB1190 on dioxane is likely dependent on the glyoxylate carboligase pathway, and is potentially enhanced by malate synthase G.

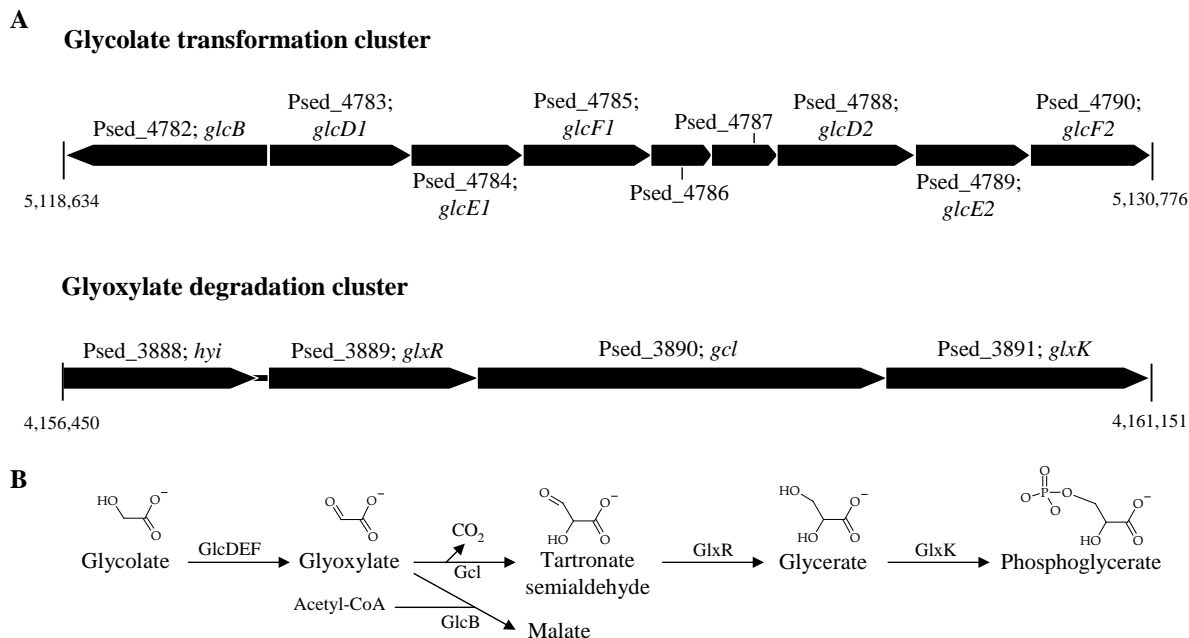


Figure 5.2. Strain CB1190 chromosomal regions of interest implicated in glycolate and glyoxylate transformations during dioxane metabolism. A) Gene clusters implicated in glycolate and glyoxylate metabolism. B) Proposed transformations of glycolate and glyoxylate catalyzed by enzymes putatively encoded by strain CB1190 gene clusters.

Since strain CB1190 can grow with dioxane, the bacterium must derive energy from the compound's transformation. The glyoxylate carbolygase pathway represents a potential route to energy generation during dioxane metabolism via glyoxylate, so we sought to confirm the activity of this pathway in dioxane-grown cells. Glyoxylate carbolygase (Gcl) activity ($2 \text{ glyoxylate} \rightarrow \text{CO}_2 + \text{tartronate semialdehyde}$, Fig. 5.2) was tested in cell-free extracts prepared from strain CB1190 cells. As Fig. 5.3A shows, pre-growth on dioxane resulted in significantly greater consumption of glyoxylate than pre-growth on pyruvate. Glyoxylate disappearance was accompanied by a stoichiometric production of CO_2 , a specific product of the glyoxylate carbolygase reaction. Heating of the dioxane cell-free extracts for 10 min at 80°C reduced CO_2 production to that observed with the buffer treatment (data not shown). No CO_2 was detected in dioxane cell-free extracts when glyoxylate was omitted.

To verify the putative roles of strain CB1190 genes in the dioxane-induced glyoxylate carbolygase pathway, we cloned and over-expressed Psed_3890 (putative glyoxylate carbolygase gene) and Psed_3889 (putative tartronate semialdehyde reductase gene, Fig. 5.2) in *Rhodococcus jostii* strain RHA1 using the thiostrepton-inducible plasmid pTip-QC2 (102). As expected, glyoxylate was consumed and CO_2 produced only in cell-free extracts from Psed_3890-expressing strain RHA1, or when these extracts were mixed with extracts from Psed_3889-expressing RHA1 (Fig. 5.3B). The mixture of cell-free extracts from RHA1/pTip-Psed_3890 and RHA1/pTip-Psed_3889 also produced a small amount of glycerate ($0.71 \pm 0.23 \mu\text{mol per vial}$). No glycerate was produced with cell-free extracts from either construct alone, or with cell-free extracts from strain RHA1 with empty pTip-QC2. The amount of glycerate detected with the mix of RHA1/pTip-Psed_3890 and RHA1/pTip-Psed_3889 cell-free extracts represented $\sim 7\%$ of that expected if tartronate semialdehyde was stoichiometrically transformed to glycerate (Fig. 5.2) as

a dead-end product. Tartronate semialdehyde was not detectable with the HPLC method used, so it is unclear whether the small amount of glycerate detected was due to only a minor transformation of tartronate semialdehyde or due to the further transformation of the generated glycerate to unspecified products. Regardless, these results support the annotation of Psed_3890 as a glyoxylate carboligase gene and Psed_3889 as a tartronate semialdehyde reductase gene.

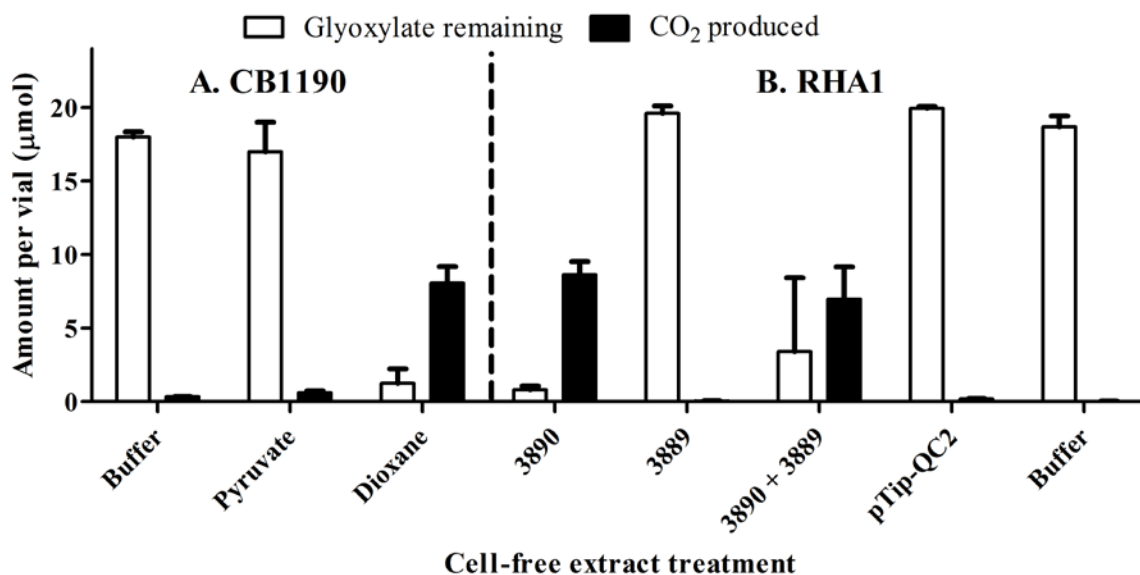


Figure 5.3. Glyoxylate carboligase activity in strain CB1190 and strain RHA1/pTip cell-free extracts. **A)** Phosphate buffer or cell-free extracts from pyruvate- or dioxane-grown strain CB1190 cells were amended with 10 mM glyoxylate and then glyoxylate remaining and CO₂ production were determined. **B)** The same assay was performed with cell-free extracts prepared from strain RHA1 cells with pTip-Psed_3890 (3890; putative glyoxylate carboligase), pTip-3889 (3889; putative tartronate semialdehyde reductase), a mix of the two (3890 + 3889), pTip-QC2 (empty vector), or buffer.

Amino acid isotopomer analysis to identify routes for dioxane carbon assimilation

The glyoxylate carboligase pathway ultimately leads to the assimilation of carbon as three-carbon metabolites in glycolysis as 3-phosphoglycerate and later pyruvate (49). In order to obtain further support for the hypothesis that carbon from dioxane enters central metabolism via the glyoxylate carboligase pathway as three-carbon compounds, we performed isotopomer amino acid analysis with ¹³C-fully-labeled dioxane. Strain CB1190 was grown either with ¹³C-dioxane as the sole carbon source or with ¹³C-dioxane and elevated (3%) CO₂. With ¹³C-dioxane as the sole carbon source, all of the detected amino acids were heavily labeled, which confirmed, as expected, that strain CB1190 can directly use dioxane as a sole carbon source.

Pyruvate-derived alanine, valine, phenylalanine and tyrosine were predominantly ¹³C-labeled when strain CB1190 was grown with ¹³C-dioxane as the sole carbon source, and the addition of elevated unlabeled CO₂ did not affect these labeling patterns. These results are in contrast to aspartate, threonine, serine and glycine; for these, a small proportion of unlabeled carbon was integrated when the cells were grown with ¹³C-dioxane as the sole carbon source, and this proportion increased greatly when strain CB1190 was grown with ¹³C-dioxane and elevated unlabeled CO₂. These results indicate that carbon from CO₂ was assimilated during the synthesis of these amino acids (Fig. 5.4).

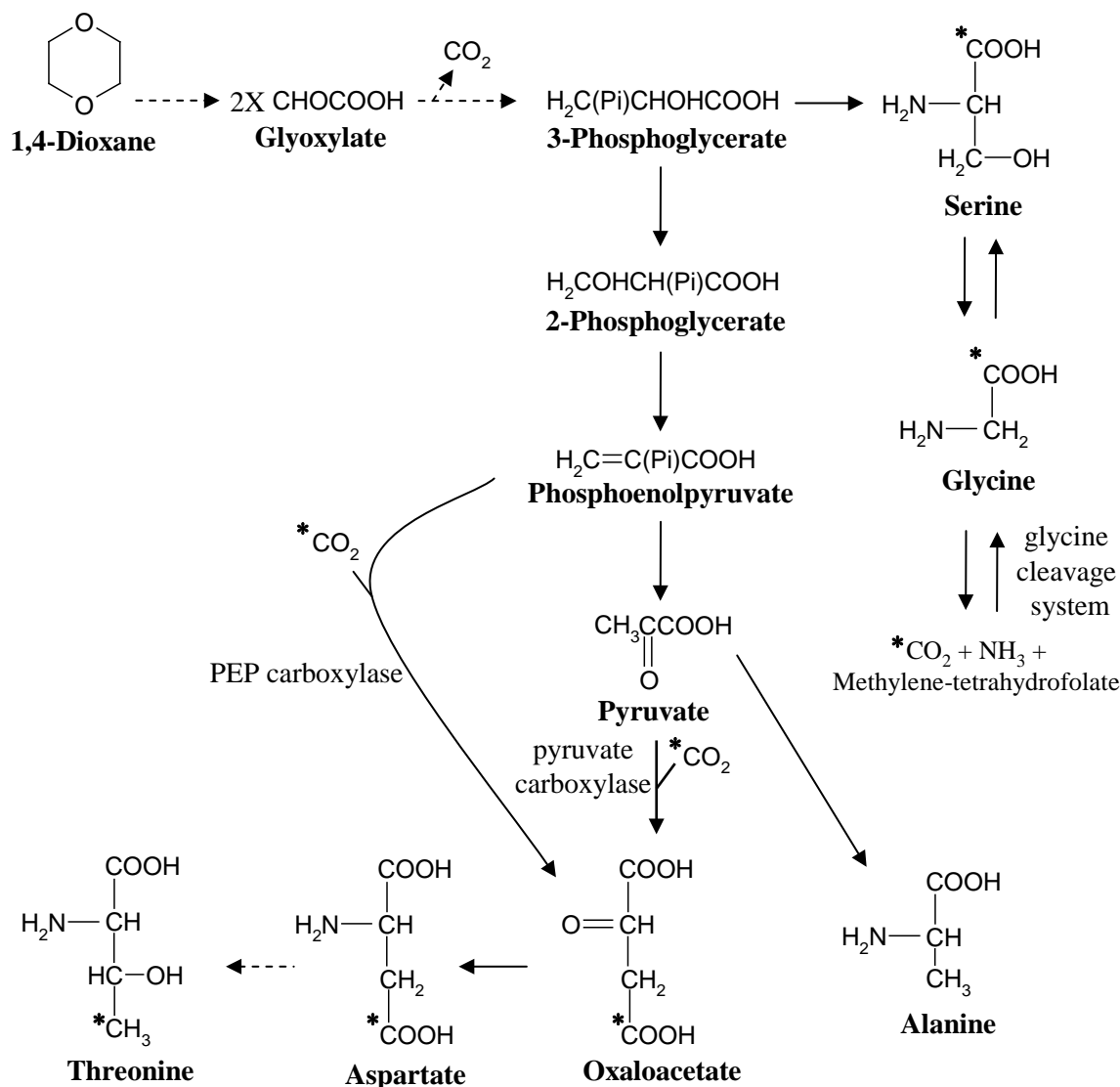


Figure 5.4. Proposed assimilation of unlabeled carbon during growth of strain CB1190 with uniformly ¹³C-labeled dioxane. Unlabeled carbons are indicated with an asterisk (*); all other carbons are considered ¹³C. The dashed arrow indicates a multi-step reaction. The amino acids valine, phenylalanine and tyrosine were omitted for simplicity.

For the oxaloacetate-derived aspartate and threonine, the unlabeled carbon was present at a non-first carbon position under both experimental conditions. These results indicate that the anaplerotic pathway transforming either phosphoenolpyruvate (PEP) or pyruvate to oxaloacetate through the addition of CO₂ (69, 171) was active during dioxane metabolism (Fig. 5.4). Strain CB1190 has homologous genes for both PEP carboxylase (Psed_6164) and pyruvate carboxylase (Psed_3032), which catalyze these anaplerotic reactions. The enzyme that catalyzes the reverse of the PEP-utilizing process, PEP carboxykinase, is putatively encoded by Psed_6619, and this gene was down-regulated during growth with both dioxane and glycolate, relative to pyruvate. Taken together, these results indicate that with dioxane, carbon flows predominantly from the direction of PEP towards the TCA cycle, and support the proposed role of the three-carbon

compound-generating glyoxylate carboligase pathway in dioxane metabolism. In contrast, when grown with pyruvate, strain CB1190 needs to generate PEP and further gluconeogenic compounds necessary for amino acid synthesis, which explains the relative upregulation of the gene (Psed_6619) encoding the PEP-generating PEP carboxykinase.

The inconsistency between the labeling patterns of 3-phosphoglycerate-derived serine and glycine and pyruvate-derived alanine, valine, phenylalanine and tyrosine is noteworthy, since 3-phosphoglycerate is a precursor of pyruvate in glycolysis. An explanation for this inconsistency is an alternative pathway of serine and glycine biosynthesis, specifically, by the reverse activity of the glycine cleavage system (GCS). The GCS converts glycine and tetrahydrofolate to CO_2 , NH_3 and methylene-tetrahydrofolate (129), and it is putatively encoded by the strain CB1190 genome (Chapter 7). The GCS is reversible (67), and reverse GCS activity in the presence of elevated unlabeled CO_2 would result in increased unlabeled carbon in the glycine pool, and consequently, in the serine pool as well (Fig. 5.4). Thus, during dioxane metabolism, it is possible that serine and glycine are synthesized from both the glyoxylate carboligase pathway via 3-phosphoglycerate and the reverse GCS.

Glyoxylate could contribute to the serine pool in yet another way. The strain CB1190 glyoxylate carboligase cluster has a gene (*hyi*) for hydroxypyruvate isomerase, which catalyzes the isomerization of hydroxypyruvate to tartronate semialdehyde (6) and which is up-regulated in strain CB1190 during growth with dioxane (Tables 5.2 and 5.3). Hydroxypyruvate can be used as a precursor for serine biosynthesis by the activity of transaminases (133). However, no genes with significant homology to known amino acid-glyoxylate aminotransferase or amino acid-hydroxypyruvate transaminase genes were up-regulated in strain CB1190 during growth with dioxane or glycolate, so the role of hydroxypyruvate isomerase in dioxane metabolism remains unclear.

Energy generation in strain CB1190 dioxane metabolism

Based on the current work, we present an updated proposed dioxane degradation pathway for strain CB1190, linking dioxane metabolites to central metabolism (Fig. 5.1). The ability of strain CB1190 to grow on dioxane necessitates energy generation, and the employment of the glyoxylate carboligase pathway during dioxane degradation conceptually provides the basis for energy production, since this pathway results in the pyruvate precursors that can then be metabolized to yield NADH and ATP (73). An analysis of the reducing equivalents consumed or produced in the individual steps proposed for dioxane metabolism in Fig. 5.1 reveals a maximum net yield of 6 NADH and one GTP per dioxane. Assuming an equivalence of 3 NADH per ATP, this means the theoretical maximum yield of ATP generated per dioxane is 19. Mahendra and Alvarez-Cohen (89) previously reported a yield of 0.09 g protein per g dioxane for strain CB1190, so assuming that protein constitutes 50% of dry cell weight and integrating this cell yield with the theoretical maximum yield of ATP from dioxane, this results in a Y_{ATP} (9) of 0.83 g dry cell weight per mol ATP. In comparison, *Lactobacillus casei* growing with glucose (128) and *Actinomyces* strains growing with lactate (162) had Y_{ATP} values of 24 and 10 g dry cell weight per mol ATP, respectively.

This work represents the first insights into the genetic basis for microbial dioxane metabolism and provides a more complete picture of the pathway leading to carbon assimilation and energy generation during growth on dioxane. It will be useful to determine if dioxane supports growth through a similar pathway in the other reported dioxane-metabolizing microorganisms, including *Pseudonocardia benzenivorans* strain B5 (89), *Mycobacterium* sp. PH-06 (66), *C. sinensis* (101) and *Rhodococcus ruber* strain 219 (13).

6. Transcriptomics of THF degradation and investigation of monooxygenase activities in *Pseudonocardia dioxanivorans* strain CB1190

This is a summary of work published in:

1. Sales, C. M. 2012. Functional genomics of bacterial degradation of the emerging contaminants 1,4-dioxane and N-nitrosodimethylamine (NDMA). Ph.D. thesis. University of California, Berkeley.
2. Additional results presented here are in publications in preparation.

Introduction

A number of studies have reported the ability of pure and mixed cultures of bacteria and fungi to degrade dioxane aerobically (13, 19, 66, 89, 101, 112, 125, 127, 141, 161, 180) and one study reported anaerobic degradation (139). The most common mechanism for aerobic dioxane biodegradation is co-metabolism, by which another substrate is used to provide carbon and energy to the microorganism. However, a handful of bacteria, including *Rhodococcus ruber* 219 (13), *Mycobacterium* sp. PH-06 (66), *Pseudonocardia dioxanivorans* CB1190 (112), *Pseudonocardia benzenivorans* B5 (89) and the fungus *Cordyceps sinensis* (101), are capable of growth on dioxane as their sole carbon and energy source. In contrast, metabolism of the cyclic ether tetrahydrofuran (THF) as a sole carbon and energy source among bacteria is more common. Many of actinobacteria can utilize THF for growth (13, 68, 112, 154, 161) as can some fungal isolates (102, 115, 141).

The initial step in the aerobic biodegradation of both dioxane and THF is catalyzed by a monooxygenase reaction. Metabolic pathways for THF degradation have been proposed for *R. ruber* 219 (13), *Pseudonocardia tetrahydrofuranoxydans* K1 (154), and *Graphium* sp. (141). Common among these metabolic pathways is the initial oxidation of THF to 2-hydroxytetrahydrofuran and the formation of succinate as the downstream intermediate. No intermediates of THF degradation by *R. ruber* 219 have been detected, but the ability of *R. ruber* 219 to grow on the hypothetical intermediates γ -butyrolactone and succinate has been shown (13). In *P. tetrahydrofuranoxydans*, a gene cluster involved in the utilization of THF was cloned and sequenced (154). This gene cluster contained a four-subunit multicomponent monooxygenase, *thmADBC*, putatively involved in the oxidation of THF to 2-hydroxytetrahydrofuran; a succinate semialdehyde dehydrogenase, *sad*, responsible for the conversion of succinic semialdehyde to succinate; and an aldehyde dehydrogenase, *aldH*. In the fungus *Graphium* sp., a cytochrome P450 monooxygenase enzyme was suggested to catalyze the initial oxidation of THF. The only intermediate of THF degradation detected in *Graphium* sp., or any other strain, has been γ -butyrolactone (141). In both *P. tetrahydrofuranoxydans* and *Graphium* sp., the conversion of 2-hydroxytetrahydrofuran to γ -butyrolactone is proposed to be catalyzed by an alcohol dehydrogenase. Furthermore, the conversion of γ -butyrolactone to 4-hydroxybutyrate is proposed to be an abiotic process and the subsequent reaction of 4-hydroxybutyrate to succinic semialdehyde is catalyzed by an alcohol dehydrogenase. Homologues of the *thm* cluster from *P. tetrahydrofuranoxydans* have been found in *Pseudonocardia* sp. ENV478 (161), *Rhodococcus* sp. YYL (173), and *P. dioxanivorans* CB1190 (132).

In Chapters 4 and 5 we used genomics, transcriptomics and biochemical assays to gain insights into dioxane metabolism by strain CB1190. These studies identified a gene cluster encoding a putative multi-component monooxygenase homologous to the THF-degradation-associated *thmADBC* gene cluster from *P. tetrahydrofuranooxydans* K1 (154) and strain ENV478 (93). Both dioxane and THF degradation pathways are initiated by a monooxygenase, as discussed above, but in strain CB1190 the genetic basis for THF degradation has not yet been examined. The purpose of the work in this chapter is therefore two-fold: 1) determine the genes involved in THF metabolism in strain CB1190, and 2) investigate the functional role of the monooxygenase gene cluster *thmADBC* in THF and dioxane degradation and in the transformation of the dioxane metabolism intermediate HEAA.

Methods

Chemicals

The potassium salt of β -hydroxyethoxyacetic acid (HEAA) was prepared by CanSyn Chemical Corporation (Toronto, ON).

Culture conditions

Cells of strain CB1190 were grown in ammonium mineral salts (AMS) liquid medium (112). The growth substrate dioxane, THF, isopropanol, succinate, or glucose were added to the culture medium to achieve a final concentration of 5 mM. HEAA was added to cultures as a growth substrate with a final concentration of 1.5 mM. Cultures were incubated aerobically while shaking at 30°C.

Cell harvesting and RNA isolation for transcriptional studies

Strain CB1190 cells were harvested and RNA purified for transcriptomic microarray analysis and quantitative reverse transcriptase PCR (qRT-PCR) experiments using the methods described in Chapter 5.

Analytical methods

Dioxane and THF removal was monitored by injection of liquid samples onto a Varian 3400 gas chromatograph with a flame ionization detector and an 1% AT-1000 on a carbograph1 packed column (Grace, Columbia, MD). Samples were analyzed isothermally at 170°C, and the injector and detector temperatures were 230°C and 250°C, respectively.

The presence of HEAA in samples was measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a hydrophilic interaction chromatography (HILIC) column. Samples were filtered with 0.2 μ m syringe filters to remove cells. Standards and samples were measured on an Agilent Technologies 1200 Series LC system equipped with an Agilent Technologies Zorbax HILIC Plus column (4.6 mm x 100 mm, 3.5 μ m) coupled to an Agilent Technologies 6410 tandem triple quadrupole (QQQ) mass spectrometer (Santa Clara, CA). The LC solvents used were Solvent A: aqueous buffer and Solvent B: acetonitrile. The aqueous buffer was 0.45 μ m filtered, 10 mM ammonium formate in water with formic acid to adjust acidity (pH 5). The flow rate was 0.5 mL/min. The gradient was: t = 0 min., 95% B; t = 10 min., 40% B; t = 13 min., 95% B; t = 15 min., 95% B. The injection volume for samples and standards was 10 μ L. Column effluent was introduced into the electrospray chamber, where the electrospray ionization was set to 3000 V in negative mode (ESI-). Nitrogen was used as the nebulizing gas at 30 psi, 325°C, and a gas flow of 11 L/min. Multiple reaction monitoring

(MRM) was used to monitor the transitions from the parent ion of HEAA, m/z 119 $[C_4H_7O_4]^+$ to the major product ions of m/z 119 $[C_4H_7O_4]^+$, m/z 101 $[C_4H_5O_2]^+$, m/z 75 $[C_2H_3O_3]^+$, and m/z 31 $[CH_3O]^+$ during a run. The collision gas used for MRM was argon, with the collision energy of 6 V and a fragmentor of 70 V for all MRM transitions. The summation of the peak areas for all MRM transitions was used for quantification of standards and samples.

Transcriptomics microarray analysis

Five μ g of total RNA was used as starting material for each microarray chip, which were processed using the protocol described in Chapter 5. All microarray data analyses were performed in the R statistical programming environment (www.r-project.org) using packages available from Bioconductor version 2.9 (www.bioconductor.org) (Gentleman *et al.*, 2004). Hybridization signal intensities for probe sets were calculated using the “rma” function from the “affy” package (41). The “rma” function implements the computation of the RMA expression measure (57), which summarizes the probe intensities for each probe set, by background adjustment, quantile normalization (15), and the median polish linear fitting procedure.

Identification of differentially expressed genes was accomplished using the “limma” package (143), which implements a linear models approach to analyzing designed microarray experiments. This approach first required fitting the microarray data to a linear model, specified by a design matrix that indicates which RNA samples have been applied to each array, using the function “lmFit”. Since the expression data was from Affymetrix GeneChips, the linear modeling was the same as one-way analysis of variance (ANOVA). The second step in this approach involved specifying the comparisons to be made between groups of arrays in a contrast matrix, such as contrasting all treatments (dioxane, glycolate, THF, and succinate) to the control (pyruvate) and direct comparison between treatments (*e.g.*, dioxane to glycolate or THF to succinate). This contrast matrix was applied to the linear fitted microarray data using the function “contrast.fit”. Estimated log-fold changes (\log_2FC) and statistics used to assess differential expression between contrasts were determined using an empirical Bayes method, implemented in the function “eBayes”. The P values returned from linear modeling were adjusted to correct for multiple hypothesis testing by applying the Benjamini and Hochberg procedure (11). This procedure allowed for the control of the false discovery rate (FDR) to 1% (*i.e.*, by only considering contrasts with an adjusted p -value less than 0.01). Genes that met the criteria of $FDR < 1\%$ and a $\log_2FC \geq |1|$ were considered differentially expressed.

Quantitative RT-PCR analyses

Total RNA was isolated from strain CB1190 cells after 8 hours of exposure to either 5 mM of dioxane, 5 mM of THF, 20% (vol./vol. of total headspace) of propane, or 5 mM of glucose (control). Cells for induction analysis were grown in AMS medium with 5 mM of glucose and harvested by filtration. The harvested cells were washed with and resuspended in AMS medium. The resuspended cells were aliquoted so each treatment and control condition was tested in triplicate. The isolated RNA was used to synthesize cDNA using the TaqMan reverse transcription kit (Applied Biosystems). TaqMan chemistry was used for qPCR reactions targeting *thmA* (Psed_6976), *prmA* (Psed_0639), *tpi* (Psed_3417), *thiC* (Psed_6168), and *rpoD* (Psed_0376). The genes *tpi*, *thiC*, and *rpoD* were determined by the geNorm method (163) to be stably transcribed across all treatments, and were thus used as internal references. Each qPCR reaction, which was run in triplicate for each biological replicate on an Applied Biosystems StepOne Plus real-time PCR system, consisted of 1X Fast Universal Mix (Applied Biosystems), 2 μ L of diluted cDNA, each primer at 0.5 mM, and the probe at 145 nM. The cycling conditions

were: 95°C for 20 s, then 40 cycles of 95°C for 1 s followed by annealing for 20 s. The efficiency of each qPCR assay was determined with a serial dilution of cDNA derived from a dioxane treatment replicate. Gene expression was normalized using the method of Vandesompele *et al.* (163).

***thm* gene cluster transcriptional analysis**

Cells for RT-PCR analysis of the THF monooxygenase genes *thmADBC* were grown on dioxane, isopropanol, or THF and harvested by centrifugation. RNA was isolated using the RNeasy kit. RT-PCR was performed using the Qiagen One Step RT-PCR kit. The primers used for amplification of each fragment were: fragment A, *thmA_For1* 5'-ATGACTGCCCCACCGATGAAG-3' and *thmA_Rev* 5'-AACGGTAGTAGTCGTCATTACACC-3'; fragment B, *thmA_For2* 5'-CCTATAAGCGAGTGGAATAC-3' and *thmD_Rev* 5'-CGATGAGTAACGCCGAACGACTC-3'; fragment C, *thmD_For* 5'-GCTGCGCCAGATGTCAGAGG-3' and *thmC_Rev1* 5'-CCACATAATCATAAGCGACGTCG-3'; and fragment D, *thmB_For* 5'-CTCATGAGCGCGAGATCGAG-3', *thmC_Rev2* 5'-CCACATAATCATAAGCGACGTCG-3'. RT-PCR products are visualized on an electrophoresis gel.

Cloning and expression of strain CB1190 THF monooxygenase genes in *R. jostii* strain RHA1

The 4.3 Kb fragment of strain CB1190 genes *Psed_6976-6979* was initially PCR-amplified and cloned into the plasmid pK18 (121). The *Psed_6976-6979* fragment encoding for the THF monooxygenase genes *thmADBC* was then subcloned into plasmid pTip-QC2 (102). The fragments was amplified with forward primer 5'-AAGGAGATATACATATGACTGCCCCACCGAT-GAA-3' and reverse primer 5'-GTATGCGGCCCGCCATGGAATTCTACGACTCAGA-GTTGATCAGCTCGAT-3', where the gene-binding sequence is underlined and the NdeI and XhoI restriction sites, respectively, are in bold. Each 100 µL PCR reaction consisted of 1X HF Buffer, 200 nM dNTPs, 500 nM each primer, 3% DMSO, 2 units Phusion Hot Start II DNA polymerase (Finnzymes) and 10 ng of plasmid pT7-7/*thfmo* DNA. Thermocycling conditions were as follows: 98°C for 3 min, then 30 cycles of (98°C for 20 s, 72°C for 6 min), then 72°C for 10 min. The appropriate amplicon was then gel-purified.

Plasmid pTip-QC2 was linearized with NdeI and EcoRI (New England Biolabs), gel purified, and then the plasmid and PCR insert were ligated at a 1:3 (plasmid:insert) ratio at 16°C overnight with T4 DNA ligase (New England Biolabs). Electrocompetent *E. coli* DH5α was transformed via electroporation with 1 µL of ligation mix. Ampicillin-resistant colonies were screened for the appropriate construct, which was named pTip-*thfmo*.

Fifty ng of purified plasmid pTip-*thfmo* or empty vector pTip-QC2 were used to transform electrocompetent *Rhodococcus jostii* strain RHA1 (strain RHA1) according to the method of Kalscheuer *et al.* (61), except that cells were not pre-incubated prior to electroporation, and the electroporation parameters were: 12 kV/cm, 800 Ω; and 25 µF.

Single colonies of strain RHA1 containing the plasmid pTip constructs were used to inoculate nutrient broth amended with chloramphenicol and grown at 30°C with shaking for 48 h. Two liter flasks containing 0.5 L of nutrient broth amended with chloramphenicol and 1% each of sucrose and glycine were inoculated with 10 mL of 48 h culture and incubated at 30°C with shaking. When the cultures had reached OD600 of 0.8 (~21 h), thiostrepton dissolved in

DMSO was added to a final concentration of 1 µg/mL and the cultures were incubated for a further 26 h. Cells were transferred to centrifuge bottles (Beckman Coulter, Brea, CA), cooled on ice, pelleted by centrifugation, and then cell pellets were washed twice with ice-cold buffer (20 mM sodium phosphate, pH 7.0, 0.05% Tween-80). The cells were finally resuspended in ice-cold buffer, aliquoted to 1.5 mL microcentrifuge tubes, and then stored at -80°C until use in enzyme assays.

Aliquots of strain RHA1 carrying plasmid pTip constructs were thawed and resuspended in 1 mL of cold buffer. The transformation of dioxane and THF and production of HEAA by cell suspensions of RHA1/pTip-thfmo or RHA1/pTip-QC2 was tested in 2 mL microcentrifuge tubes containing 1 mL of buffer and cell suspension. Each triplicate tube received 100 µL of cell suspension, and was amended with dioxane (2.5 mM), THF (4.0 mM) or HEAA (10 mM) from aqueous stocks. Buffer controls (no cell suspension) vials were prepared in parallel. Vials were incubated at 30°C with horizontal shaking at. At several timepoints, 200 µL were removed and analyzed for the disappearance of the amended compound. GC-FID (for dioxane and THF) and LC-MS/MS (HEAA) was used for analyzing samples.

HEAA transformation inhibition assay

For assays to test inhibition of HEAA degradation by acetylene, strain CB1190 was grown with dioxane and harvested by filtration. Cells were washed on the filter with phosphate buffer to remove residual dioxane, and then scraped cells were resuspended in phosphate buffer. One half of the cell suspension was exposed briefly to 5% (vol./vol.) acetylene in the headspace for 10 minutes while shaking at 30°C. N₂ gas was then used to remove residual acetylene. The non-acetylene-exposed half of the cell suspension was treated identically, without the addition of acetylene. Finally, HEAA was amended to the cell suspensions and abiotic controls and the concentration of HEAA in each suspension was monitored over time.

Results and Discussion

Transcriptomics of *P. dioxanivorans* strain CB1190 growth on THF and succinate

Succinate is an intermediate of aerobic THF degradation in the bacterium *P. tetrahydrofuranoxydans* (154) and the fungus *Graphium* sp. (141). Therefore, in order to determine the genes involved in the degradation of THF in strain CB1190, the transcriptomes during THF and succinate growth were also compared directly and relative to pyruvate. The transcriptomes of THF- and succinate-grown strain CB1190 were determined by microarray analysis. When the THF- and succinate-grown transcriptomes were individually contrasted against the pyruvate-grown transcriptome, 717 and 580 genes were differentially expressed, respectively. Growth on THF versus pyruvate led to the up-regulation of 224 genes and the down-regulation of 493 genes. Growth on succinate versus pyruvate led to the up-regulation of 377 genes and the down-regulation of 203 genes.

A total of 830 genes were differentially expressed on THF versus succinate, of which, 220 were up-regulated and 610 were down-regulated. A total of 92 genes are up-regulated on THF and not on succinate, both relative to pyruvate (Table 6.1).

Table 6.1 Genes up-regulated on THF but not succinate, relative to pyruvate.

Up-regulated genes have a $\log_2\text{FC} \geq 1$, FDR (adjusted p -value) < 0.01 (not shown). [Genes highlighted in **Red** are not differentially expressed $\log_2\text{FC}$ and FDR < 0.01 when comparing the transcriptomes of THF- and succinate-grown strain CB1190 directly.]

Gene	Protein	THF vs. Pyruvate $\log_2\text{FC}$	Succinate vs. Pyruvate $\log_2\text{FC}$
Psed_0094	YCII-related	2.46	0.77
Psed_0095	RNA polymerase sigma factor, sigma-70 family	1.52	0.21
Psed_0129	protein of unknown function DUF59	1.62	0.67
Psed_0130	amidohydrolase 2	2.17	0.83
Psed_0131	Alcohol dehydrogenase GroES domain protein	1.99	0.60
Psed_0288	S-(hydroxymethyl)glutathione dehydrogenase	1.76	0.18
Psed_0355	hypothetical protein	1.26	-0.41
Psed_0357	helix-turn-helix domain protein	1.08	0.06
Psed_0382	hypothetical protein	1.16	0.84
Psed_0554	hypothetical protein	2.06	0.59
Psed_0934	YCII-related	2.58	-0.16
Psed_1301	Redoxin domain protein	1.03	-0.05
Psed_1303	Heavy metal transport/detoxification protein	1.05	-0.72
Psed_1584	6-phosphofructokinase	1.05	-0.49
Psed_1967	major facilitator superfamily MFS_1	1.85	0.09
Psed_1968	hypothetical protein	2.11	0.27
Psed_2321	ABC-type sugar transport system periplasmic component-like	1.22	-0.29
Psed_3256	hypothetical protein	3.34	0.99
Psed_3442	protein of unknown function DUF939	1.05	0.74
Psed_3564	Aromatic-amino-acid transaminase	1.66	0.47
Psed_3652	hypothetical protein	1.52	-0.25
Psed_3669	hypothetical protein	2.28	0.66
Psed_3986	Peptidase M23	1.05	0.54
Psed_4083	hypothetical protein	1.23	0.80
Psed_4236	hypothetical protein	1.03	-0.01
Psed_4268	membrane protein of unknown function UCP014873	2.06	0.12
Psed_4321	hypothetical protein	1.55	-0.63
Psed_4322	Dihydrolipoyl dehydrogenase	2.21	-0.42
Psed_4699	protein of unknown function DUF1365	1.62	0.83
Psed_4702	protein of unknown function DUF1295	3.20	0.52
Psed_4753	nitrogen regulatory protein P-II	1.15	0.06
Psed_4754	nitrogen regulatory protein P-II	1.34	0.41
Psed_4937	hypothetical protein	1.06	-0.61
Psed_4938	regulatory protein MarR	2.83	-0.16
Psed_4939	Lipase	6.70	-0.18

Psed_4940	fumarylacetoacetate (FAA) hydrolase	5.21	-0.01
Psed_4941	Formyl-CoA transferase	2.78	0.02
Psed_5082	YVTN beta-propeller repeat-containing protein	1.20	0.23
Psed_5176	hypothetical protein	1.61	0.66
Psed_5224	hypothetical protein	1.19	-0.55
Psed_5245	regulatory protein TetR	1.13	-0.15
Psed_5246	uncharacterized peroxidase-related enzyme	1.64	0.01
Psed_5247	hypothetical protein	2.10	-0.13
Psed_5248	Nitroreductase	4.37	-0.04
Psed_5249	NADPH:quinone reductase	3.45	-0.25
Psed_5250	major facilitator superfamily MFS_1	1.04	0.10
Psed_5736	Transglycosylase-like domain protein	2.30	0.94
Psed_5747	hypothetical protein	1.83	0.53
Psed_5748	hypothetical protein	1.10	0.58
Psed_5749	hypothetical protein	1.31	0.70
Psed_5752	hypothetical protein	1.04	0.32
Psed_5753	hypothetical protein	1.73	0.29
Psed_5755	hypothetical protein	1.21	0.66
Psed_5756	hypothetical protein	1.16	0.61
Psed_5757	hypothetical protein	1.79	0.18
Psed_5758	hypothetical protein	2.27	0.60
Psed_5759	hypothetical protein	2.39	0.45
Psed_5760	hypothetical protein	2.22	0.26
Psed_5761	hypothetical protein	2.02	0.61
Psed_5762	hypothetical protein	1.63	0.33
Psed_5763	hypothetical protein	1.35	0.35
Psed_5764	hypothetical protein	1.31	0.22
Psed_5765	hypothetical protein	1.01	0.21
Psed_5766	hypothetical protein	1.43	0.53
Psed_5767	hypothetical protein	1.50	0.14
Psed_5768	hypothetical protein	1.43	0.56
Psed_5769	hypothetical protein	1.55	0.19
Psed_5770	hypothetical protein	1.29	0.29
Psed_5778	cell division FtsK/SpoIIIE	1.33	0.12
Psed_5779	hypothetical protein	2.01	0.08
Psed_5780	hypothetical protein	1.37	0.78
Psed_5781	hypothetical protein	1.29	0.81
Psed_6160	heavy metal translocating P-type ATPase	1.24	0.92
Psed_6249	Ribulose biphosphate carboxylase large chain	1.15	-0.20
Psed_6665	NLP/P60 protein	2.19	0.27
Psed_6780	hypothetical protein	1.57	-0.10
Psed_6806	hypothetical protein	1.31	-0.12
Psed_6807	hypothetical protein	2.79	-0.11

Psed_6812	hypothetical protein	1.37	0.92
Psed_6847	hypothetical protein	1.11	0.08
Psed_6857	protein of unknown function DUF156	1.03	0.72
Psed_6879	S-adenosylmethionine synthase	1.82	-0.23
Psed_6970	D-lactate dehydrogenase (cytochrome)	2.71	-0.20
Psed_6971	Hydroxyacid-oxoacid transhydrogenase	2.17	0.44
Psed_6972	GntR domain protein	1.66	0.31
Psed_6974	Ethyl tert-butyl ether degradation EthD	1.82	0.92
Psed_6975	Betaine-aldehyde dehydrogenase	1.63	0.73
Psed_6977	Ferredoxin--NAD(+) reductase	2.48	0.70
Psed_6978	methane/phenol/toluene hydroxylase	1.59	0.54
Psed_6979	monooxygenase component MmoB/DmpM	2.23	0.68
Psed_6981	Aldehyde Dehydrogenase	3.35	0.88
Psed_6982	Mn ²⁺ /Fe ²⁺ transporter, NRAMP family	3.46	0.25

The *thm* gene cluster (Psed_6974 to Psed_6982) located on plasmid pPSED02 is found in the set of genes up-regulated on THF but not succinate, relative to pyruvate. The most highly up-regulated gene in the *thm* cluster is the Mn²⁺/Fe²⁺ transporter. Another gene (Psed_6971) located just upstream of the *thm* cluster on plasmid pPSED02 is also up-regulated. The gene Psed_6971 is annotated to encode for a hydroxyacid-oxoacid transhydrogenase, which is known to convert γ -hydroxybutyrate to succinic semialdehyde, using α -ketoglutarate as a hydrogen acceptor (Kaufmann *et al*, 1988). Other oxoacids, including oxaloacetate and α -ketoapitate can also be used as a hydrogen acceptor but with lower activity. Psed_6970 is also up-regulated on THF versus succinate, relative to pyruvate, and encodes for a lactate dehydrogenase, a homologue of γ -hydroxybutyrate dehydrogenase, which converts γ -hydroxybutyrate to succinic semialdehyde using NADH.

Transcriptomic microarray analysis of THF- and succinate-grown strain CB1190 cells agreed with the Northern blot analysis reported for *P. tetrahydrofuranoxydans* that showed genes in the *thm* cluster to be expressed during THF utilization (154). Although a probe set for the THF monooxygenase oxygenase component α -subunit *thmA* is not on the microarray, RT-qPCR shows that THF induces gene expression of *thmA* by 10.2-fold ($\log_2\text{FC} = 3.35$), relative to glucose. The only disagreement between the Northern blot analysis performed with *P. tetrahydrofuranoxydans* and the transcriptional analyses performed with strain CB1190 is that the aldehyde dehydrogenase gene (Psed_6981), which has been proposed to be involved in the conversion of 4-hydroxybutyraldehyde to 4-hydroxybutyrate in *P. tetrahydrofuranoxydans* and *Graphium* sp. (141), is up-regulated during growth on THF in strain CB1190 and not in *P. tetrahydrofuranoxydans*. An alcohol dehydrogenase had been previously proposed to catalyze the conversion of 2-hydroxytetrahydrofuran to γ -butyrolactone (141). Among the genes only up-regulated on THF relative to pyruvate is one annotated as an alcohol dehydrogenase GroES domain protein (Psed_0131). Although the activity of this alcohol dehydrogenase needs to be confirmed, this is the first evidence of induced alcohol dehydrogenase gene expression during THF degradation. The conversion of 4-hydroxybutyrate to succinate semialdehyde is proposed to be catalyzed by the product of at least one of two differentially expressed genes on the plasmid pPSED02, a lactate dehydrogenase (Psed_6970) or hydroxyacid-oxoacid transhydrogenase (Psed_6971).

Therefore, based on transcriptomic microarray analysis, the proposed THF degradation pathway starts with (i) the hydroxylation of THF to 2-hydroxytetrahydrofuran by the THF monooxygenase (*thmADBC*), (ii) conversion of 2-hydroxytetrahydrofuran to γ -butyrolactone by the alcohol dehydrogenase (Psed_0131), (iii) abiotic hydrolysis of γ -butyrolactone to produce 4-hydroxybutyrate, (iv) conversion of 4-hydroxybutyrate to succinate semialdehyde via either transhydrogenase (Psed_6971) or dehydrogenase (Psed_6970), (v) the oxidation of succinate semialdehyde to succinate by succinate semialdehyde dehydrogenase (*sad*), (vi) the succinate finally enters the tricarboxylate cycle. The 4-hydroxybutyrate could potentially be polymerized to poly-beta-hydroxybutyrate (PHB) since two poly-beta-hydroxybutyrate polymerase genes (Psed_2979 and Psed_2990) exist on the strain CB1190 chromosome.

Induction of *thmA* gene expression by dioxane and THF

The microarray used in this study is missing a number of genes from the finished strain CB1190 genome and does not include sequences determined to be pseudogenes. The sequence Psed_6976 was designated as a pseudogene and was subsequently left off the microarray. However, sequence analysis of the *thm* gene cluster determined that the Psed_6976 shares 98.9% amino acid identity, 94% nucleotide identity, and the same gene order with the tetrahydrofuran monooxygenase α -subunit sequence, *thmA*, characterized in *P. tetrahydrofuranooxydans*. RT-PCR analysis of the THF monooxygenase genes *thmADBC* (Fig. 6.1) indicated that all of the genes in this cluster were transcribed during growth on isopropanol, dioxane, or THF. Amplification of fragment A demonstrated that Psed_6976 is not a pseudogene but an expressible *thmA* gene. The weak amplification of fragments B and D that span *thmAD* and *thmDBC*, respectively, suggested that the reductase component (*thmD*) is weakly co-transcribed with its neighboring open-reading frames. The presence of fragment D shows that two genes *thmB* and *thmC* were co-transcribed.

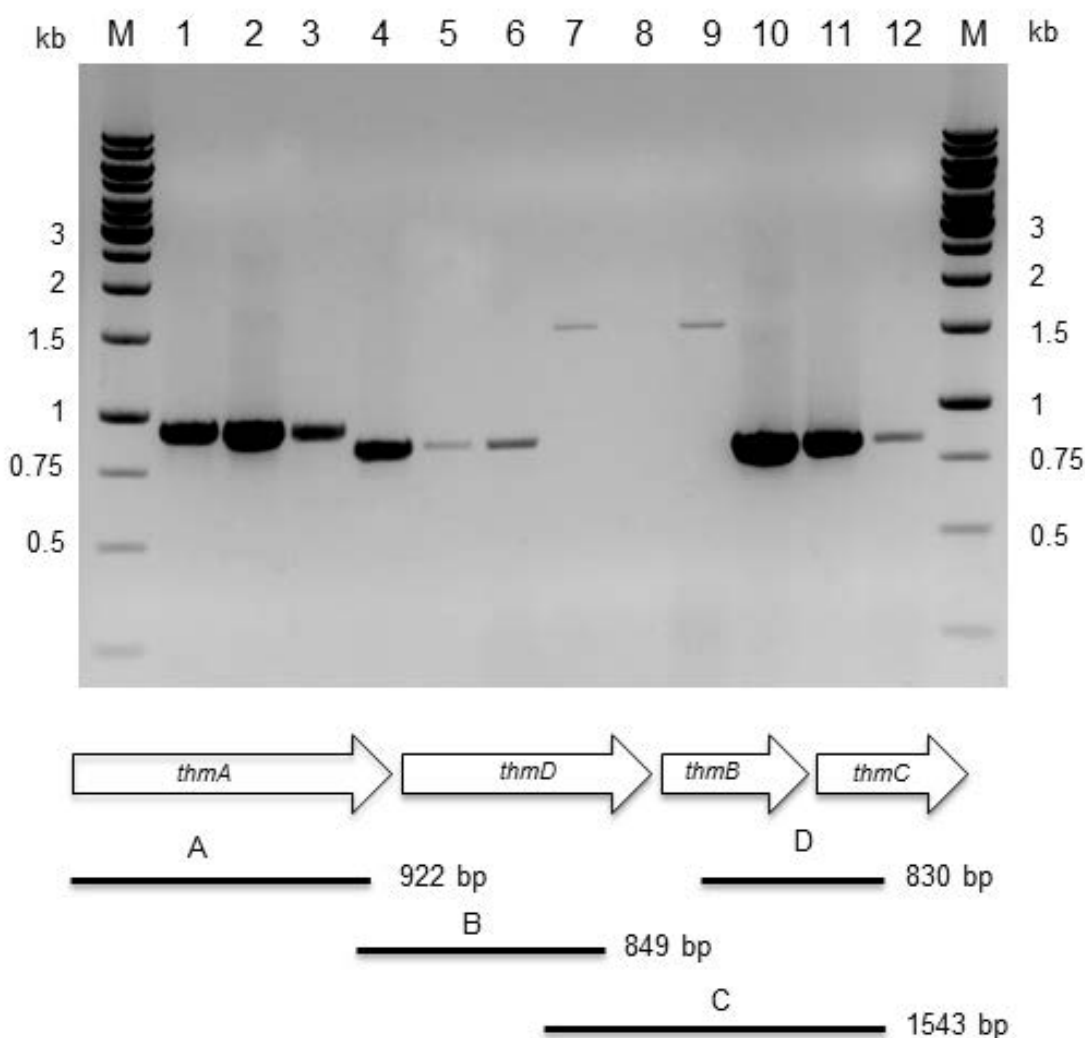


Figure 6.1. RT-PCR analysis of THF monooxygenase gene expression.

Lanes 1-3, primers *thmA_For1* and *thmA_Rev* were used to amplify fragment A. Lanes 4-6, primers *thmA_For2* and *thmD_Rev* were used to amplify fragment B. Lanes 7-9, primers *thmD_For* and *thmC_Rev1* were used to amplify fragment C. Lanes 10-12, primers *thmB_For* and *thmC_Rev2* were used to amplify fragment D. RNA used in reactions shown in lanes 1, 4, 7, and 10 was purified from isopropanol-grown cells, RNA used in reactions shown in lanes 2, 5, 8, and 11 was purified from dioxane-grown cells, and RNA used in reactions shown in lanes 3, 6, 9, and 12 was purified from THF-grown cells. Lanes M, molecular weight markers. No products were detected in reactions lacking reverse transcriptase.

qRT-PCR analysis had been used to quantify the up-regulation of *thmA* (Psed_6976) during growth on dioxane (Chapter 5). Here, qRT-PCR was used to quantify the induction of *thmA* gene expression after 8 hours of exposure to dioxane, THF, or propane. In cells exposed to dioxane, THF, and propane, *thmA* was induced 24, 10, and 6 fold, respectively, relative to glucose (Fig. 6.2). A similar analysis was performed for *prmA* (Psed_0629), encoding propane monooxygenase α -subunit. Exposure to dioxane and THF did not lead to significant differential

expression of *prmA*, with a differential-fold expression of 2 and 1, respectively. However, propane significantly induced the *prmA* gene (64-fold relative to glucose).

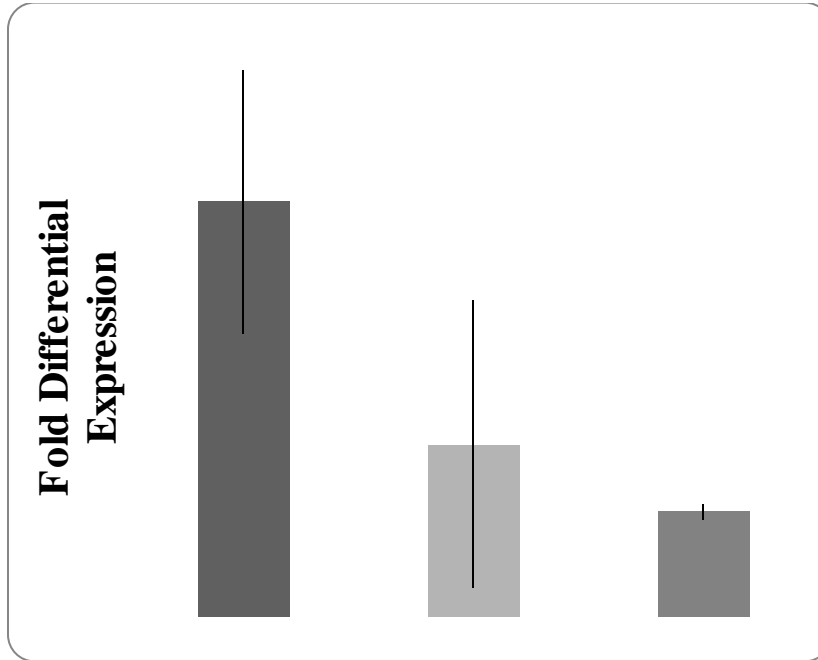


Figure 6.2. Induction of *thmA* gene expression in glucose-grown strain CB1190.

Glucose-grown cells were exposed to dioxane, THF, or propane and after 8 hours of exposure, RNA was extracted and transcripts were quantified using qRT-PCR. Fold-differential expression values were calculated relative to the expression of cells exposed to glucose (control) and were normalized to housekeeping genes.

Functional activity of *thmADBC*-expressing *R. jostii* RHA1 clones

The strain CB1190 gene cluster *thmADBC* encoding the putative THF multicomponent monooxygenase was cloned and expressed in *Rhodococcus jostii* strain RHA1 from the *Rhodococcus* expression vector pTip-QC2 (RHA1/pTip-thfmo). The RHA1/pTip-thfmo clones were grown in nutrient broth and expression was induced by thiostrepton. Monooxygenase activity was tested with dioxane. RHA1/pTip-thfmo transformed 2.5 mM of dioxane into HEAA within 3 days (Fig. 6.3). The generation of HEAA was stoichiometric with dioxane removal. Dioxane was not removed in either abiotic controls or with the empty vector pTip-QC2-containing RHA1 clone RHA1/pTip-QC2. The compound HEAA was detectable at extremely low concentrations in the abiotic and pTip-QC2 controls, indicating that HEAA is a minor impurity in the dioxane stock from the manufacturer.

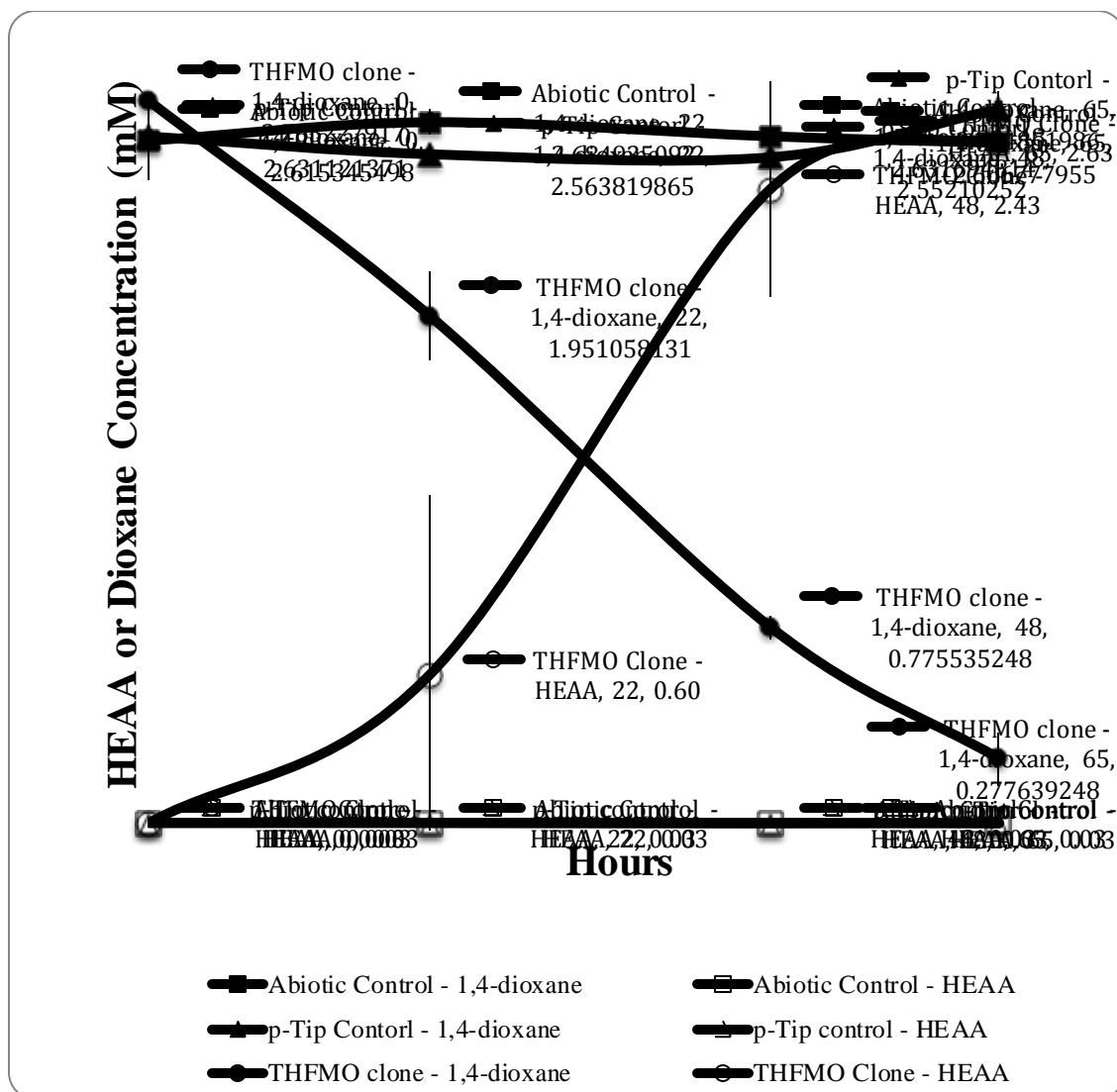


Figure 6.3. Transformation of dioxane to HEAA by heterologous strain CB1190 monoxygenase. RHA1/pTip-thfmo, RHA1/pTip-QC2 empty vector control or abiotic control was amended with dioxane and dioxane disappearance and HEAA production was monitored. Error bars represent the range of duplicate reactions.

The RHA1/pTip-thfmo expression clone also removed THF, with 4 mM of THF disappearing within 3 days (Fig. 6.4). THF loss in the abiotic and pTip-QC2 controls was minimal and was attributable to the volatility of THF.

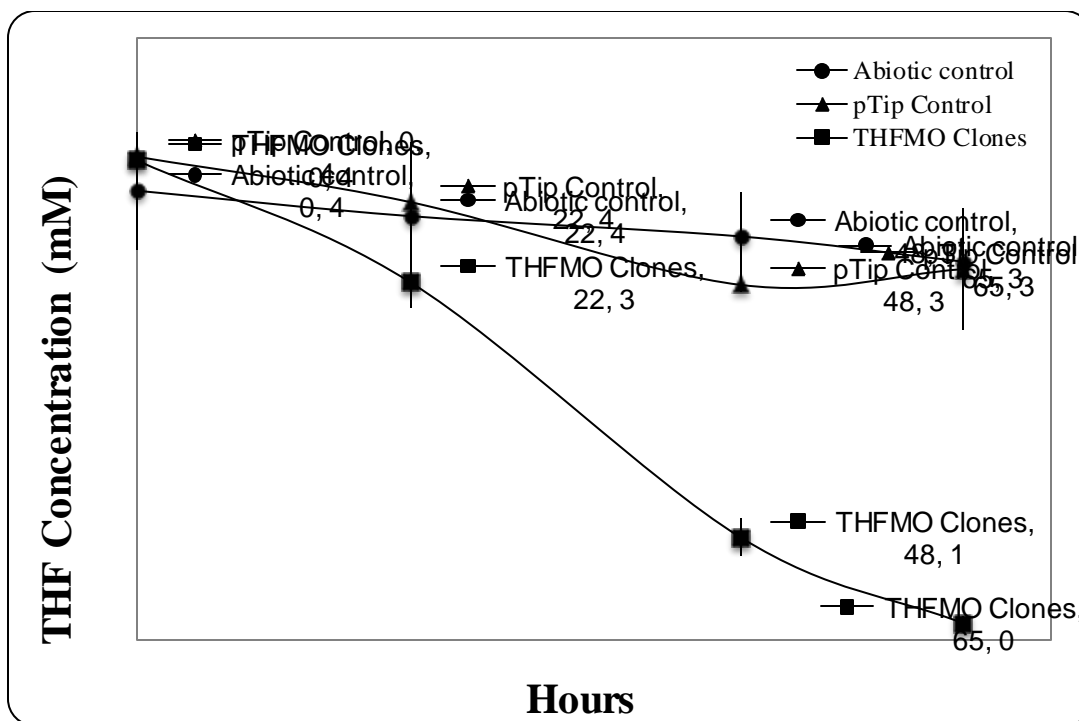


Figure 6.4. Removal of THF by heterologous strain CB1190 monooxygenase.

RHA1/pTip-thfmo, RHA1/pTip-QC2 empty vector control or abiotic control was amended with THF and THF removal was monitored. Error bars represent the range of duplicate reactions.

These experiments confirm that the strain CB1190 genes Psed_6976-6979 encode a multi-component monooxygenase that is active with both dioxane and THF. This is the first functional expression of a monooxygenase from a dioxane-metabolizing microorganism and from any *Pseudonocardia* sp.

Growth of *P. dioxanivorans* strain CB1190 on HEAA

Growth of strain CB1190 on identified intermediates of dioxane, such as ethylene glycol, glycolate, glycoaldehyde, glyoxal, glyoxylic acid, oxalic acid, and formic acid have been confirmed (91). However, growth on 2-hydroxy-1,4-dioxane, 1,4-dioxane-2-one (dioxanone), HEAA, and 1,2-dihydroxyethoxy-2-hydroxyethoxy acetic acid have not been determined. We obtained HEAA in its salt form in order to test the growth of strain CB1190 on HEAA as sole carbon and energy source. Strain CB1190 grew with HEAA as verified by visual inspection (data not shown) and HEAA removal (Fig. 6.5). Starting with a 1:500 inoculation of strain CB1190 cells previously growing on dioxane, complete removal was accomplished within 11 days.

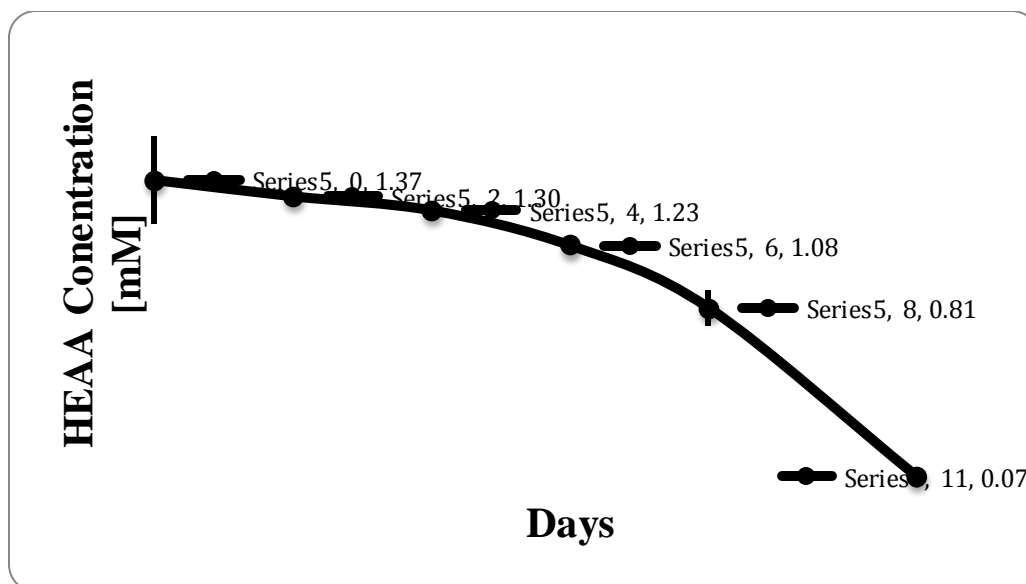


Figure 6.5. Degradation of HEAA by strain CB1190.
Cells of strain CB1190 were grown in AMS medium amended with HEAA (n = 5).

Testing the effect of acetylene exposure on HEAA degradation

Mahendra *et al.* (91) proposed that in the dioxane metabolic pathway in strain CB1190 a monooxygenase is also involved in the hydroxylation of HEAA and that this monooxygenase is likely the same enzyme that catalyzes the initial hydroxylation of dioxane. However, after incubating the dioxane-degrading RHA1/pTip-thfmo with chemically synthesized HEAA, no HEAA removal was observed in 72 h (data not shown).

Since brief exposure to acetylene gas causes irreversible inhibition in some monooxygenases (*e.g.*, (24, 123, 136, 147)) and specifically inhibits co-metabolic and metabolic degradation of dioxane (89, 91), dioxane-grown strain CB1190 cells were briefly exposed to acetylene gas to test if transformation of HEAA is indeed catalyzed by a monooxygenase reaction. As Fig. 6.6 shows, brief exposure had minimal effects on HEAA removal, with 1.25 mM of HEAA removed within 90 h.

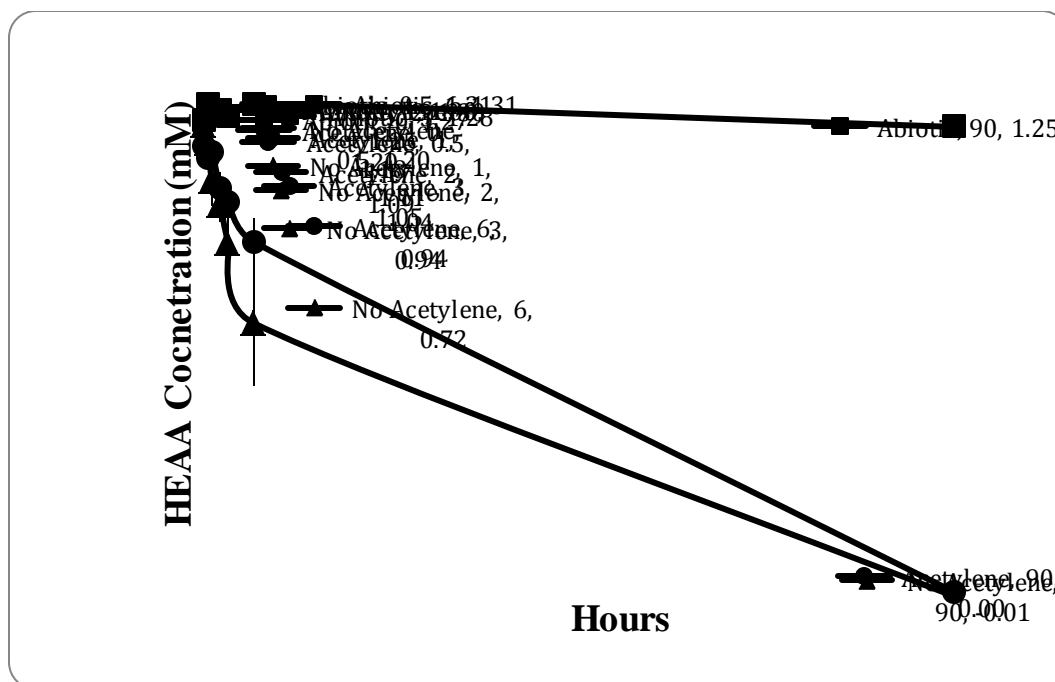


Figure 6.6. Test of acetylene on HEAA removal by strain CB1190. Dioxane-grown cells were exposed to acetylene and then HEAA removal was monitored. Squares, abiotic samples; Triangles, acetylene-exposed cells; Circles, no-acetylene-exposure cells. Error bars represent the range of triplicate samples.

These results contradict the proposal by Mahendra *et al.* (91) that the same monooxygenase acts on both dioxane and HEAA. If a monooxygenase does in fact catalyze the transformation of HEAA in strain CB1190, this enzyme is different from that encoded by Psed_6976-6979, and is unlike many other monooxygenases in that it is resistant to irreversible inhibition by acetylene.

Conclusion

This study identified for the first time genes used by *Pseudonocardia dioxanivorans* strain CB1190 to grow with THF. The monooxygenase involved in the initial hydroxylation of both THF and dioxane was confirmed by cloning and heterologous expression in *R. jostii* strain RHA1. Experimental evidence with these RHA1 clones and with the monooxygenase-inhibiting gas acetylene contradicts previous suggestions that the same monooxygenase acts on both dioxane and on the dioxane metabolic intermediate HEAA – it appears that HEAA is transformed via a still unknown mechanism. While substantial advances were achieved in understanding the dioxane metabolic pathway both in Chapter 5 and in the current Chapter, further study is required to understand how dioxane C4 intermediates are transformed to C2 compounds.

7. Using metabolomics analyses to assist genome annotation of *Pseudonocardia dioxanivorans* strain CB1190

1. Results presented in this section are included in publications in preparation.

Introduction

Strain CB1190 and genome annotation

Pseudonocardia dioxanivorans CB1190 was previously isolated from industrial sludge contaminated with dioxane (112) and it is one of the first bacterium reported to be capable of growth on dioxane and the first member of the genus *Pseudonocardia* for which there is an annotated genome sequence (132). In addition to dioxane, strain CB1190 is capable of using a broad spectrum of organic compounds as its sole carbon and energy source, including other ethers (e.g., tetrahydrofuran, diethyl ether, and butyl methyl ether), various alcohols and aromatic compounds (e.g., benzene and toluene) (90, 112). It can also grow autotrophically with CO₂. The ability to use a wide variety of carbon sources is not surprising given the large genome (7.3 Mb) (132). However, the presence of multiple apparent carbon utilization pathways for each individual compound, multiple gene copies of each enzyme in one particular pathway, and often missing key enzymes in pathways create ambiguities in the genome annotation. For example, strain CB1190 seems to bear three CO₂ fixation pathways (a Calvin-Benson-Bassham (CBB) pathway, a reverse citric acid cycle, and a Wood-Ljungdahl pathway), of which none is complete. Therefore, a genome independent method is necessary to evaluate the strain CB1190 genome annotation in order to more precisely connect the genomic information and strain CB1190's physiological traits. In this work we chose to use C2 metabolic pathways as test cases for verifying genome annotation.

Ethylmalonyl-CoA pathway and glyoxylate cycle pathway

The transformation of dioxane by strain CB1190 eventually leads to the production of two glyoxylate molecules, which are assimilated via the glyoxylate carboligase pathway (see Chapter 5). Coincidentally, glyoxylate is also an important intermediate involved in aerobic bacterial assimilation of acetyl-CoA (mostly synthesized from C1 and C2 compounds –e.g., acetate) because more complex building blocks, such as C3 compounds (e.g., pyruvate) can only be synthesized from acetyl-CoA. However, since pyruvate synthase (EC 1.2.7.1) is incapable of synthesizing pyruvate directly from acetyl-CoA under aerobic condition, alternative pathways have to be used to complete this task in aerobic bacteria (2, 5). Only two aerobic acetyl-CoA assimilation pathways are currently known: the glyoxylate cycle and the ethylmalonyl-CoA pathway (used in acetate metabolism) (2, 5, 34-37, 105, 118, 179). Both pathways convert acetyl-CoA to malate by condensation with glyoxylate, and from there pyruvate, oxaloacetate, and α -ketoglutarate are made by established steps in carbon metabolism. Despite similar substrates and products, these two pathways are differentiated in a couple of ways: the glyoxylate cycle generates reducing equivalents as NADH while ethylmalonyl-CoA pathway consumes reducing equivalents as NADPH, and carbon fixation happens in ethylmalonyl-CoA pathway but not in glyoxylate cycle (Fig 7.1) (2). A recent survey of all sequenced bacterial genomes revealed that while more than 30% of these genomes encode for the glyoxylate cycle, only 7% of them encode for the ethylmalonyl-CoA pathway, and 1% encode both pathways (2).

Strain CB1190's genome seems to encode both pathways, but ambiguities exist in the genome annotation. For example, the isocitrate lyase (EC 4.1.3.1) in the glyoxylate cycle pathway is diverged from conventional isocitrate lyase at the amino acid sequence level, and key enzymes acetoacetyl-CoA reductase (EC 1.1.1.36), mesaconyl-CoA hydratase (EC 4.2.1.-) and malate-CoA ligase (EC 6.2.1.9) in the ethylmalonyl CoA pathway are missing (131).

Glycine cleavage system and glycine-glyoxylate pathway

As opposed to acetate metabolism through the ethylmalonyl CoA pathway, glycine assimilation in aerobic bacteria represents a non-acetyl-CoA-related C2 compound assimilation strategy which is generally achieved through the glycine cleavage system (GCS). Here, one glycine molecule is cleaved into a CO₂ and a methyl group as methyl-THF (tetrahydrofolate), which in turn reacts with a second glycine molecule to synthesize a serine. From the serine, pyruvate and other central pathway metabolites can be synthesized and used to produce energy and building blocks.

Additionally, some bacteria can generate glyoxylate from glycine through the direct oxidation of glycine by D-amino-acid oxidase (EC 1.4.3.19). The strain CB1190 genome encodes a D-amino-acid oxidase homologue. Since glyoxylate metabolism through the glyoxylate carboligase pathway is established in CB1190 (Chapter 5), glycine may also be metabolized via glyoxylate as an important intermediate.

Research questions, hypotheses and important results

The work presented here had two motivations. First, the ambiguities in the genome annotation must be clarified so that the genetic information can be more effectively leveraged. For example, an unambiguous genome annotation is needed to construct precise genome-scale models that allow us to predict behaviors of strain CB1190 in dioxane bioremediation processes. Therefore, verifying the annotation of genes involved in the assimilation of C2 compounds represents a test case for resolving genome annotation ambiguities. Second, from a practical standpoint, given that the metabolism of environmental C2 compounds could interfere with dioxane metabolism (by virtue of their sharing the glyoxylate carboligase pathway), it is important to understand if the highlighted C2 compound assimilation pathways are in fact active in strain CB1190. To these ends, in this study we analyzed acetate and glycine metabolic pathways in strain CB1190 at the functional and genetic levels.

Materials and Methods

Chemicals

[1-¹³C]sodium pyruvate, [1-¹³C]sodium acetate, [2-¹³C]sodium acetate and [¹³C]sodium bicarbonate (99% purity) were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA).

Bacterial strain and culture conditions

Strain CB1190 was routinely cultivated in ammonium mineral salts (ASM) medium (112). Replicate 250 mL glass bottles with screw-caps were filled with AMS medium and were amended with the appropriate carbon source. Dioxane-grown strain CB1190 was used as inoculants with 1% inoculation in order to start culture lines with respective carbon source. To unambiguously track the carbon transformations and incorporations in the metabolic network, the tracer experiments with respective isotopic carbon sources (10 – 30 mM) were conducted in

duplicates which were [1-¹³C]sodium pyruvate, [1-¹³C]sodium acetate, [2-¹³C]sodium acetate, and glycine (50 mM) with [¹³C]sodium bicarbonate. In order to avoid unlabeled carbon introduced from inoculation, strain CB1190 was subcultured with 1% inoculum in ASM medium with isotopic carbon source three times before being harvested for isotopomer analysis.

Analytical methods

Analysis of amino acid isotopomers was performed as previously described (149, 184).

RNA extraction for transcription studies

Cells were harvested when approximately half of the added substrate had been consumed. Cells from replicate bottles were collected by filtration, were scraped from the filters with a sterile scalpel, and were transferred to 2 mL screw-top microcentrifuge tubes containing 1 g 100- μ m-diameter zirconia beads (Biospec Products, Bartlesville, OK). These tubes were stored at -80°C until use. Nucleic acids were extracted using a modified version of the phenol method described previously (60). DNA was removed by treatment with DNase I, and RNA was finally purified with an RNeasy Kit (Qiagen).

Reverse transcript-PCR

cDNA was synthesized from RNA using the TaqMan reverse transcription kit (Applied Biosystems, Foster City, CA) with random hexamers according the manufacturer's protocol. Synthesized cDNA was used as template in each subsequent PCR reaction. All PCR reactions were performed with gDNA as positive control, milliQ-H₂O as no template negative control and negative controls using cDNA synthesized from target gene absent conditions. 0.8% agarose gel was used to separate the PCR products and evaluate the results.

Cell-free extracts and enzyme measurement

Strain CB1190 cells grown on respective carbon source were collected by filtration, scraped and stored in 2 mL microcentrifuge tubes at -80°C. Zirconia beads were added, and the cells were resuspended in 100 mM Tris•HCl (pH 8.2) buffer. Cells were broken by bead-beating, and then debris was collected by centrifugation. The clear supernatant was transferred and represented the cell-free extract. This extract was stably stored at 4°C for a week. The protein content was determined by using the Bradford method with BSA as the standard. Citrate synthase (84), isocitrate lyase (95), acetyl-CoA C-acetyltransferase (3, 33, 35), and malate synthase (3, 33, 35) activities were tested according to published methods.

Results and Discussion

¹³C labeled pyruvate to evaluate annotation of central pathways

In order to verify active central metabolic pathways, strain CB1190 was grown using ¹³C-labeled pyruvate since pyruvate plays pivotal roles in central metabolisms. Labeling patterns of heavy carbon (¹³C) in each amino acid were verified and revealed that strain CB1190 has a complete and active central metabolic pathway, including glycolysis and citrate cycles. The ¹³C-pyruvate studies indicated the presence of an active glyoxylate cycle for quickly replenishing the citrate cycle metabolites (such as malate), which are withdrawn for building block synthesis. This was surprising, since the gene for isocitrate lyase, a key enzyme that cleaves isocitrate into glyoxylate and succinate, appeared to be absent in the strain CB1190 genome. However, a recent re-annotation of the genome data revealed the presence of a putative isocitrate lyase homologue

(Psed_4635), for which the amino acid sequence is distant from regular bacterial isocitrate lyases. This ambiguity, along with other reasons mentioned earlier in the introduction, motivated us to conduct further analyses with the aid of stable isotopic analysis.

Demonstration of acetate assimilation using stable isotopic analyses

[1-¹³C] and [2-¹³C] sodium acetate were used to probe active acetyl-CoA assimilation pathways in strain CB1190. Several representative stable isotopic labeling profiles were obtained with these substrates (Table 7.1). The [M-159] labeling profiles of alanine, serine and glycine indicated that pyruvate was not directly synthesized from acetyl-CoA. The reason is that a major portion of the [M-159] fragment from all three amino acids was not enriched with heavy carbon, which would have been highly enriched if pyruvate was directly synthesized from acetyl-CoA. This conclusion agrees with the knowledge that pyruvate synthase (EC 1.2.7.1) requires a low redox potential in order to catalyze the carboxylation reaction; therefore aerobic bacteria cannot normally use this reaction to synthesize pyruvate from acetyl-CoA. This result also indicated that either the glyoxylate or the ethylmalonyl-CoA pathway for acetyl-CoA assimilation might be functional in strain CB1190.

Table 7.1. Amino acids ¹³C stable isotopic labeling profiles used to identify acetate assimilation pathways in strain CB1190.

[1- ¹³ C] acetate grown cells [M-57] molecule							[2- ¹³ C] acetate grown cells [M-57] molecule		
Amino Acids									
<u>Mass</u>	<u>Ala</u>	<u>Ser</u>	<u>Gly</u>	<u>Asp</u>	<u>Thr</u>	<u>Met</u>	<u>Ala</u>	<u>Ser</u>	<u>Gly</u>
M+0	0.32	0.30	0.35	0.06	0.04	0.04	0.00	0.01	0.07
M+1	0.51	0.55	0.59	0.55	0.54	0.49	0.20	0.17	0.63
M+2	0.17	0.15	0.06	0.32	0.33	0.35	0.51	0.55	0.30
M+3				0.07	0.08	0.11	0.28	0.26	

In order to determine which acetyl-CoA assimilation pathway is used by strain CB1190, we chose amino acids that are directly derived from the end products of the glyoxylate cycle or ethylmalonyl-CoA pathway (malate and succinate, respectively) as indicators to interpret the results. Isotopic labeling profiles of the [M-57] fragment for the amino acids aspartate, threonine, and methionine are listed in Table 7.1. For these amino acids, ~10% of their molecules contained three heavy ¹³C carbons. Considering that the glyoxylate cycle's overall reaction can be simplified as synthesizing a molecule of malate from two molecules of acetyl-CoA, the maximum labeled carbons in malate-derived amino acids such as aspartate and threonine can only contain two or fewer heavy carbons with respect to both [1-¹³C] and [2-¹³C]sodium acetate as sole carbon source. Given 10% of the aspartate and threonine from strain CB1190 contained three heavy carbons, ethylmalonyl-CoA pathway is more likely used to assimilate acetyl-CoA by strain CB1190. However, from looking exclusively at labeling profiles, we could not rule out the possibility that glyoxylate cycle might also be active during the growth, even though the simultaneous use of divergent carbon assimilation pathways for a single substrate is rare in bacteria. Therefore, we employed enzyme assays and reverse transcript-PCR (RT-PCR) in order to provide additional evidence for determining which acetyl-CoA pathway(s) is used by strain CB1190.

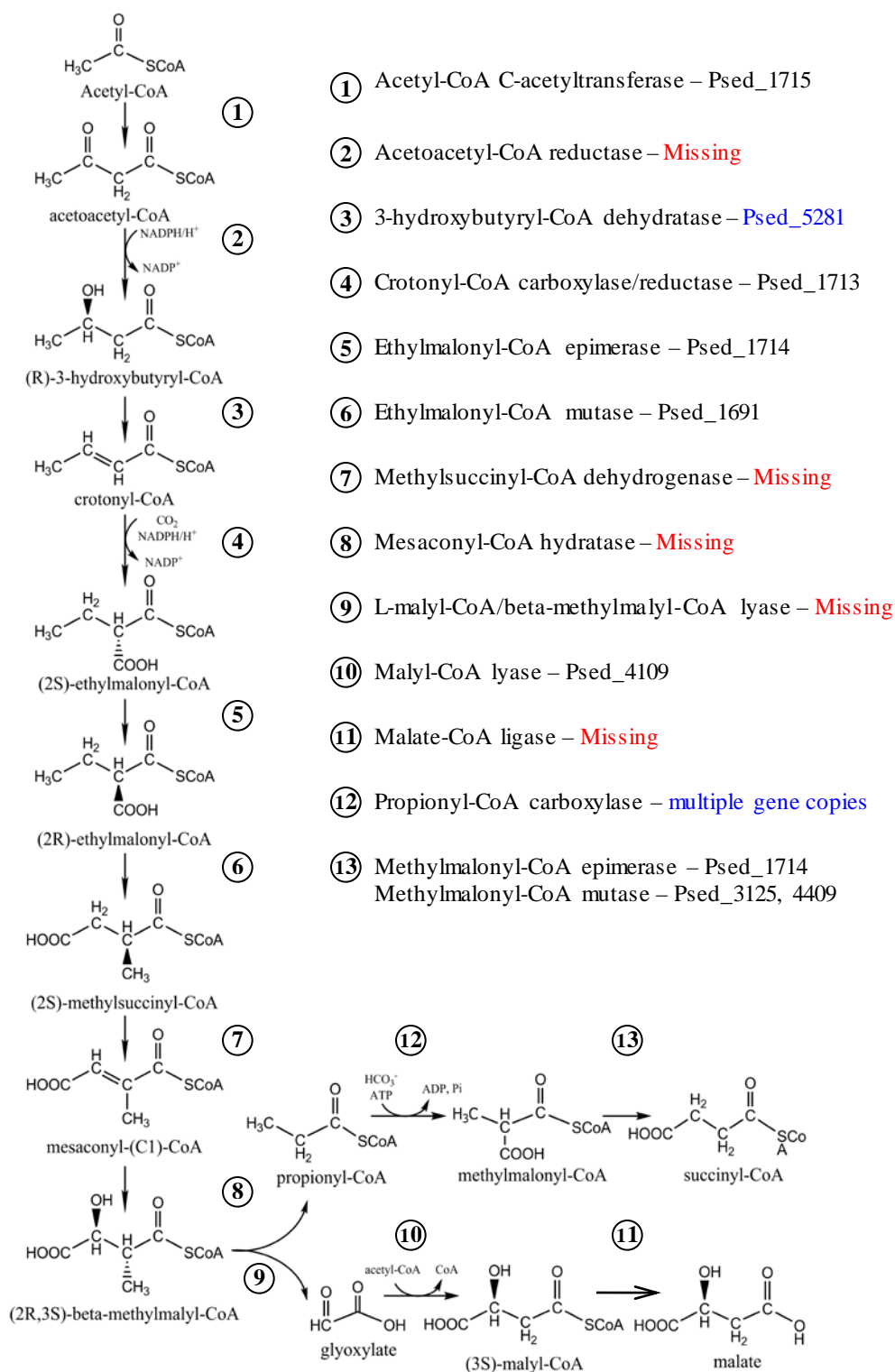


Figure 7.1. Schematic of the ethylmalonyl-CoA pathway in CB1190 with possible genes encoding the enzyme assigned to each step. Genes missing from the genome are in red.

In enzyme assay experiments, we chose isocitrate lyase from the glyoxylate cycle and acetyl-CoA C-acetyltransferase (EC 2.3.1.9) and L-malyl-CoA/beta-methylmalyl-CoA lyase (EC 4.1.3.24) from the ethylmalonyl-CoA pathway as representative enzymes to determine the activity of these pathways during strain CB1190 growth on acetate. Citrate synthase activity was used as a positive control for the activity of cell-free extracts. In isocitrate lyase assays, cell-free extracts from *E. coli* cells grown on acetate as a sole carbon and energy source were used as positive controls. In addition, extracts from glycine-grown strain CB1190 cells were used as negative controls.

Isocitrate lyase activity was not detected in acetate-grown CB1190 cell-free extracts (data not shown), which indicated that the glyoxylate cycle was probably not functional in strain CB1190 under the experimental conditions tested.

For the acetyl-CoA C-acetyltransferase and L-malyl-CoA/beta-methylmalyl-CoA lyase assays, extracts from glycine-grown strain CB1190 cells were again used as negative controls. Activity for both of these enzymes was positively detected in extracts from acetate-grown CB1190. In addition to the stable isotopic analyses and enzyme assays, we employed RT-PCR to determine whether pathway-specific genes were expressed during the growth of strain CB1190 on acetate. Total RNA from acetate-grown cells was converted to cDNA and used in PCR reactions. Strain CB1190 genomic DNA was used as a positive control for the PCR reactions, while RNA from glycine-grown cells was used as a negative control. All of the key genes selected from the ethylmalonyl-CoA pathway were detected in cDNA derived from acetate-grown cells, with the exception of the genes encoding 3-hydroxybutyryl-CoA dehydratase (EC 4.2.1.55) and malate-CoA ligase (EC 6.2.1.9).

The results demonstrate that the ethylmalonyl-CoA pathway, which was recently identified in acetate-grown *Rhodobacter sphaeroides*, is used by strain CB1190 to incorporate acetate. While glyoxylate is a metabolite in dioxane degradation, it is also one of the key intermediates in the ethylmalonyl-CoA pathway, as well as in the glyoxylate cycle. The glyoxylate cycle is commonly used by aerobic acetate utilizing microorganisms, but it appears to not be active in strain CB1190 during growth with acetate. The ethylmalonyl-CoA pathway uses three molecules of acetyl-CoA (generated directly from acetate) and two molecules of carbon dioxide to produce one molecule of malate and one molecule of succinate. Malate and succinate are then incorporated into the central tricarboxylic acid cycle (TCA cycle) to produce anaplerotic and gluconeogenic building blocks. The ^{13}C -labeling profile indicated that the two carbon dioxide molecules fixed along the ethylmalonyl-CoA pathway were predominantly from acetate oxidation. It is not yet known if the two carbon dioxide molecules could alternatively come from strain CB1190's ambient environment with elevated CO_2 level as a carbon sequestration process.

Demonstration of glycine assimilation in strain CB1190 using stable isotopic analyses

As mentioned earlier, glycine assimilation in CB1190 can also involve glyoxylate as an important intermediate, via the activity of D-amino acid oxidase. Alternatively, the glycine cleavage system (GCS) might be used to assimilate glycine. Unfortunately, with respect to the building blocks with higher carbon numbers (such as alanine or serine), both pathways give identical stable isotopic labeling patterns, regardless of what type of labeled glycine is used. This is because both pathways employ similar carbon condensation strategies to make the key C3 metabolic intermediate pyruvate from two C2 compounds (glycine in this case). Therefore, in order to differentiate these pathways, a novel ^{13}C stable isotopic analysis method was necessary (Fig. 7.2). Since enzymes involved in GCS are highly reversible (they are also involved in the synthesis and utilization of C1 compounds), we amended AMS medium with ^{13}C labeled sodium

bicarbonate along with the glycine as carbon source. If the ^{13}C labeled bicarbonate can be integrated into glycine and other amino acids, this means the GCS pathway is active under the cultural conditions provided.

Tables 2A and 2B shows that the first carbon of glycine, serine, alanine were enriched with heavy carbons, which could only come from the ^{13}C labeled bicarbonate. In addition, labeling profiles of more complex amino acids such as aspartate, threonine and methionine also confirmed the conclusion that a GCS was functional in strain CB1190. Genes potentially involved in the GCS, as identified by bioinformatic analysis, are shown in Fig. 7.3

It should be noted that since carbon reuse is a common feature in aerobic bacteria, only about 15% of alanine and serine was enriched with heavy carbon from exogenous bicarbonate. This data is not sufficient to conclude that the GCS is the sole pathway used by strain CB1190 during the glycine assimilation.

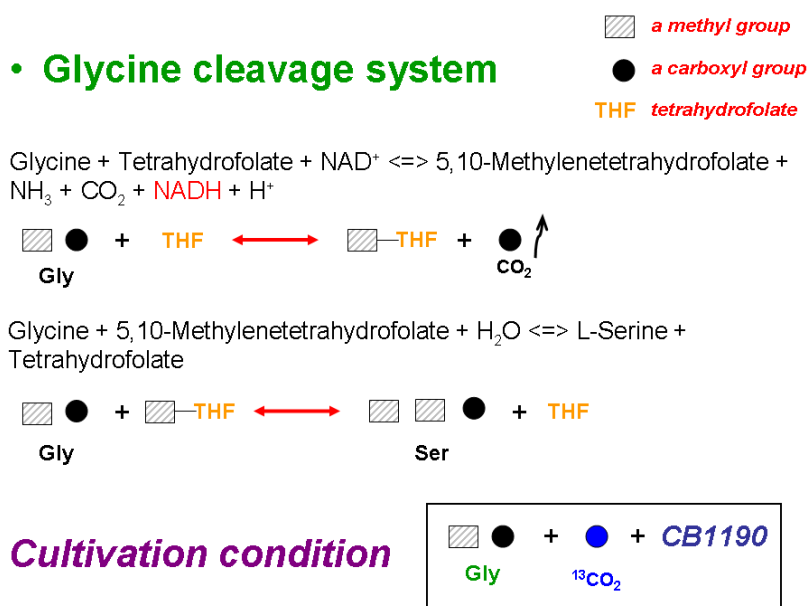


Figure 7.2. ^{13}C stable isotopic analysis method used to differentiate glycine assimilation pathways.

Table 7.2A. Amino acids ^{13}C stable isotopic labeling profiles used to identify glycine assimilation pathways in strain CB1190.

	[M-57] molecule			[M-159] molecule			[f302] molecule	
Mass	Ala	Ser	Gly	Ala	Ser	Gly	Ala	Ser
M+0	0.85	0.85	0.85		0.98	0.99	0.86	0.85
M+1	0.15	0.15	0.15	Peak	0.02	0.01	0.14	0.15
M+2	0.00	0.00	0.00	is not clear				
M+3	0.00	0.00						

Table 7.2B. Amino acids ^{13}C stable isotopic labeling profiles used to identify glycine assimilation pathways in strain CB1190.

	[M-57] molecule			[M-159] molecule		
<u>Mass</u>	<u>Asp</u>	<u>Thr</u>	<u>Met</u>	<u>Asp</u>	<u>Thr</u>	<u>Met</u>
M+0	0.59	0.56	0.56	0.70	0.69	0.68
M+1	0.35	0.38	0.37	0.29	0.31	0.31
M+2	0.06	0.06	0.07	0.00	0.00	0.00
M+3	0.00	0.00	0.00			

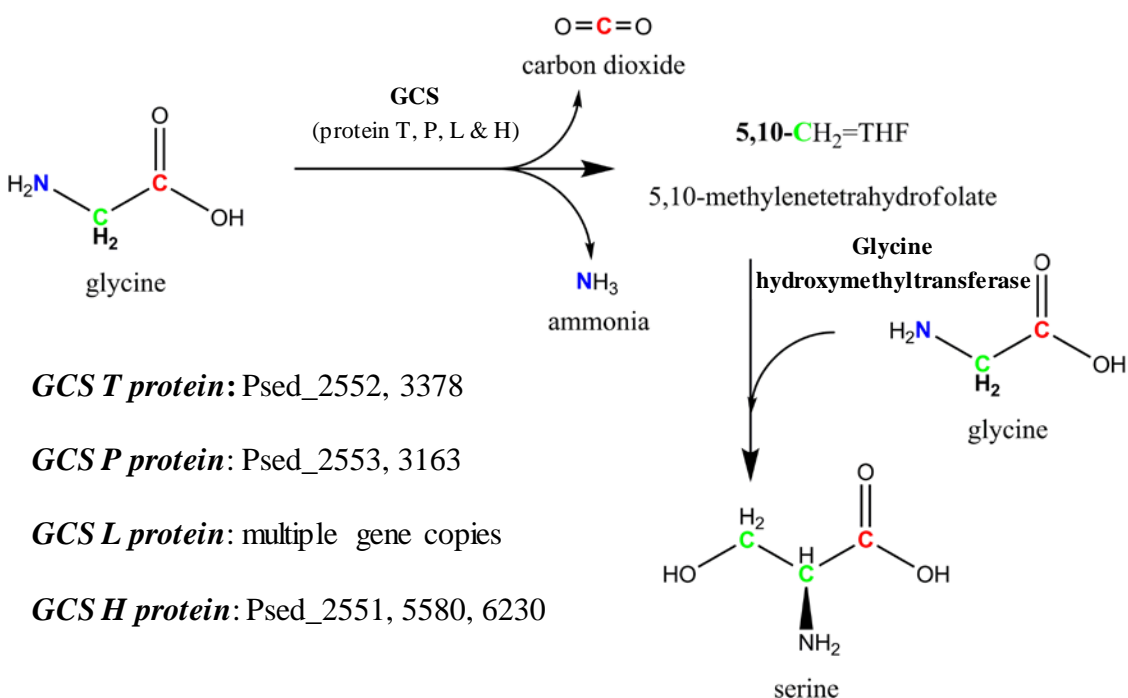


Figure 7.3. Schematic of the glycine cleavage system pathway in strain CB1190 with possible genes encoding the enzyme assigned to each step.

Conclusion

This study attempted to confirm and capitalize on the recent sequencing and annotation of the *Pseudonocardia dioxanivorans* strain CB1190 genome. We focused on the metabolism of specific C2 compounds, with the compound glyoxylate being an intermediate that unifies the metabolic pathways studied here and the dioxane metabolic pathway examined in Chapter 5. We demonstrated for the first time that the ethylmalony-CoA pathway is functional when strain CB1190 is provided sodium acetate as the sole carbon and energy source, even though several key enzyme-encoding genes are missing from the genome annotation. In addition, our results indicate that the glycine cleavage system is likely used by strain CB1190 to degrade and assimilate glycine. However, it is unclear if the glycine-glyoxylate conversion system is also

functional when glycine is provided as the sole carbon and energy source. For both pathways the detection of metabolic activities contributes to the confirmation of the annotation of genes putatively involved in these pathways.

This work, along with the application of the technique in Chapter 5, demonstrates the utility of *in vivo* ^{13}C -labeled metabolite analysis for studying fundamental metabolic pathways. This technique provides complementary and direct information for genomic and transcriptomic analysis of cellular metabolism, and would be highly applicable for confirming the interesting organic substrate degradation pathways described in Chapter 4.

8. Towards development of a genetic system for *Pseudonocardia dioxanivorans* strain CB1190

Introduction

The ability to manipulate DNA in a microorganism is desirable for two main reasons. First, interrupting genes is traditionally the pre-eminent method for linking genetic and phenotypic or metabolic traits. Second, removing, adding, or customizing a microorganism's genetic complement permits manipulation of metabolism in order to achieve a desired output. Therefore, the ability to manipulate a microorganism's DNA has impact on both understanding and engineering biology.

The capacity to use dioxane as a sole source of carbon and energy is relatively rare among tested bacterial isolates (for example, see Chapter 3). *Pseudonocardia dioxanivorans* strain CB1190 does possess this capacity, and it is the first microorganism that metabolizes dioxane for which the genome has been sequenced and annotated (see Chapter 4). Attempts to link this capacity to specific genes in order to better understand dioxane metabolism were made in Chapters 4, 5 and 6, by using bioinformatic analyses, gene expression assays (microarrays and qRT-PCR), and heterologous expression of strain CB1190 genes. However, none of these methods provided conclusive proof of the involvement and strict requirement of candidate genes in dioxane metabolism. The ability to knock out a candidate gene, observe loss of the ability to metabolically degrade dioxane, and then observe the restoration of dioxane metabolism through the addition of the candidate gene in *trans*, was desired. For this reason we endeavored to develop a genetic system for strain CB1190.

Materials and Methods

Antibiotic resistance testing of CB1190

CB1190 was grown on basal salts medium (BSM) with tetrahydrofuran (THF) as previously described (112). These cultures were used to inoculate tubes containing and the appropriate antibiotic. Tubes were incubated at 30 °C with shaking and cell growth was visually monitored and recorded daily.

Conjugative matings with *E. coli* to introduce plasmids into CB1190

Conjugative matings with *E. coli* to introduce plasmids into CB1190 were carried out on R2A plates at 30°C for 48 h. Matings using a variety of ratios of donor and recipient cells were tested. Mating mixtures were suspended in BSM and dilutions were plated on BSM plates containing appropriate antibiotics. THF was provided as the carbon source in the vapor phase.

Transposon mutagenesis of CB1190

In vitro transposition of the EZ-Tn5 <kan2> transposon into the CB1190 chromosome was carried out using the EZ-Tn5 <kan2> transpososome system (Epicentre Biotechnologies) following manufacturer's directions. Electroporation conditions were also as described in the manufacturer's instructions.

Preparation of electrocompetent cells

CB1190 actively growing in BSM containing THF was used to inoculate a fresh culture in the same medium in a 125 ml growth flask. Cells were grown at 30 °C with shaking. THF was added every other day. After 5 - 8 days, cells were harvested by centrifugation, washed three times with sterile deionized water and resuspended in a final volume of 10 ml of deionized water. Cells were stored at – 80 °C.

Electroporation of CB1190

Electrocompetent CB1190 cells were thawed on ice. After thawing, cells were mixed well using a vortex mixer and the remaining clumps of cells were allowed to settle for 2 minutes. Avoiding cell clumps, 50 µl of the cell suspension was transferred to 1.5 ml microfuge tube. Plasmid DNA (1 – 2 µl) was added and the suspension was mixed and incubated on ice for 20 minutes. The cell/DNA suspension was transferred to a chilled electroporation cuvette and placed on ice. The cells were then electroschocked and 1 ml of ice cold R2A was immediately added to the cuvette and placed on ice. The cell suspensions were transferred to a test tube and incubated at 30°C for 12 hrs without shaking. Cells were then plated onto R2A plates containing 25 µg/ml kanamycin and incubated at 30 °C. Colonies were visible within 5-10 days.

Results and Discussion

We obtained seven different plasmid vectors that function in various Gram positive bacteria: pTRW10 (erythromycin resistance [Er]), pTRKH2 (Er), pTRKL2 (Er), pSA3 (Er, tetracycline resistance [Tc], chloramphenicol resistance), pTB21 (hygromycin resistance [Hg]), pMyVec1 (kanamycin resistance) and pMyVec2 (Hg). CB1190 was found to be sensitive to 5 micrograms/ml hygromycin, kanamycin, chloramphenicol, and erythromycin, and 10 micrograms/ml tetracycline. No growth on any of the selective plates was observed for over one week, while growth on media lacking antibiotics was observed within 24 h.

E. coli strain S17-1 was transformed with each of the seven plasmids. Conjugative matings between S17-1 harboring each of the plasmids and CB1190 were attempted in order to introduce the plasmids into CB1190; however, no antibiotic resistant transformants of CB1190 were obtained in many attempts.

We tested the Transpososome system from Epicentre Biotechnologies for *in vitro* transposition of the EZ-Tn5 <kan2> transposon into the CB1190 chromosome. Successful introduction of the transposon could be used to generate random mutant libraries with marked mutations. Mutant libraries could then be screened for phenotypes of interest. We were able to verify the stable introduction of the EZ-Tn5 <kan2> transposon into the CB1190 genome; however, we only obtained 5 - 30 kanamycin-resistant CB1190 colonies from each experiment. Therefore, further optimization will be required for this to be a useful technique. However, the use of the optimized electrocompetent cells described below is expected to result in a significant improvement in efficiency.

Various electroporation conditions were tested with the seven different plasmids listed above but no antibiotic-resistant transformants of CB1190 were obtained. These results suggested that either the plasmid DNA did not enter the CB1190 cells (*i.e.*, the cells were not electrocompetent), the plasmids were not stably maintained in CB1190, or the genes required for antibiotic resistance were not being expressed. Since we had been able to show that the kanamycin resistance gene from the EZ-Tn5 <kan2> transposon could be successfully expressed

in CB1190, we decided to concentrate our efforts on pMyVec1, which carries a kanamycin resistance gene. After many attempts to optimize conditions for electroporation we were finally able to introduce pMyVec1 into CB1190 as indicated by acquisition of kanamycin resistance (Table 8.1). The important change in the preparation of electrocompetent CB1190 cells was to resuspend cells in distilled water rather than solutions containing glycerol, which is the commonly reported method. Whole cell PCR using primers to amplify the kanamycin resistance gene as well as purification of pMyVec1 from the kanamycin resistant CB1190 recombinants demonstrated that pMyVec1 had been stably introduced into strain CB1190. These results also indicate that any nucleic acid restriction systems that may be present in strain CB1190 are not fully effective at eliminating non-native DNA.

Table 8.1. Results of CB1190 electroporation with pMyVec1.

Results are averages from three independent experiments. pMyVec1 concentration was 50 ng/ul.

<u>Cuvette</u> <u>Gap width</u>				T_s	<u>CFU/ng</u>
<u>(cm)</u>	<u>kV</u>	<u>□F</u>	<u>Ω</u>	<u>(ms)</u>	<u>DNA</u>
0.1	2.5	25	200	4.9	0.62
0.1	2.5	25	400	9.5	1.80
0.1	2.5	25	800	15.7	0.95

Conclusion

This is the first report of the successful introduction of foreign DNA into *P. dioxanivorans* strain CB1190. The uptake of transpososomes creates the ability for random mutagenesis in order to discover previously unknown genes required for specific cell functions. The uptake of plasmids will make possible the functional characterization of genes in CB1190 using both directed mutagenesis and heterologous gene expression. Future needs include developing techniques for plasmid-based directed mutagenesis in order to specifically knock out genes, and determining if uptaken plasmids can act as platforms for gene expression in strain CB1190 for complementing knocked-out genes.

9. Implications of research for NDMA and dioxane bioremediation

The work presented here confirms our hypothesis that oxygenases, specifically monooxygenases, are important in the biodegradation of both NDMA and dioxane. NDMA has not yet been shown to support bacterial growth, and only a few microorganisms have been found to metabolize dioxane. Therefore, the growth of contaminant co-metabolizing microorganisms, and perhaps the induction of contaminant-transforming monooxygenases, will be dependent on the addition of a primary substrate. For NDMA degradation, propane may be an ideal primary substrate due to its low cost, lack of toxicity and demonstrated induction of NDMA-active monooxygenases. NDMA concentrations can be reduced to ng/L concentrations by co-metabolic biodegradation.

For dioxane degradation, in the absence of dioxane-metabolizing microorganisms such as *Pseudonocardia dioxanivorans* strain CB1190, a variety of primary substrates may lead to co-metabolic dioxane degradation, including methane, toluene and THF. While some of these compounds may be toxic and therefore undesirable to release into the environment, they may already be present at some dioxane-polluted sites as co-contaminants.

The potential for the inhibition of NDMA and/or dioxane biodegradation by other compounds may be a concern at contaminated sites. NDMA transformation, which is induced by the addition of propane in tested bacteria, is inhibited by high propane concentrations in a competitive, reversible manner. Therefore, an appropriate propane concentration must be added to induce NDMA-degrading microorganisms and enzymes but not completely block NDMA transformation. A benefit of co-metabolic biodegradation, however, is that upon removal by metabolism of the primary substrate the microorganism often retains transformation capacity for the secondary substrate, so NDMA may be transformed even once the added propane is consumed.

The presence of TCA and DCE at a site, which are common co-contaminants with dioxane, is a concern both because of the human toxicity of these compounds and their inhibitory effects on dioxane biodegradation. However, our results indicate the ability to co-metabolically degrade dioxane may be correlated to co-metabolic TCA and DCE transformation. Therefore, the induction of co-metabolic dioxane-degrading microorganisms with an appropriate primary substrate may in fact lead to the removal of all three co-contaminants.

Since dioxane can in fact support growth and proliferation of dioxane-degrading microorganisms, an alternative primary substrate may not be necessary to achieve dioxane removal at site. However, work presented here with strain CB1190 indicates that the co-contaminants TCA and (especially) DCE are strong inhibitors of dioxane degradation and that this bacterium is unable to transform either chlorinated solvent. Therefore, at dioxane sites with TCA or DCE present, the preferred strategy could be to stimulate co-metabolic dioxane degradation, which could be accompanied by fortuitous TCA and DCE removal.

The genome sequence of strain CB1190 indicates that this bacterium has a wide metabolic substrate range, which could be beneficial for the growth and proliferation of this dioxane-metabolizing bacterium at sites contaminated with low dioxane concentrations that poorly support growth. It is not yet known what concentration of dioxane triggers induction of enzymes necessary for dioxane transformation, but the metabolic diversity of this microorganism may make it an ideal candidate for the bioremediation of dioxane and hydrocarbon-contaminated groundwater.

10. Literature Cited

1. **Aitchison, E. W., S. L. Kelley, P. J. J. Alvarez, and J. L. Schnoor.** 2000. Phytoremediation of 1,4-dioxane by hybrid poplar trees. *Water Environ. Res.* **72**:313-321.
2. **Alber, B. E.** 2011. Biotechnological potential of the ethylmalonyl-CoA pathway. *Appl. Microbiol. Biotech.* **89**:17-25.
3. **Alber, B. E., R. Spanheimer, C. Ebenau-Jehle, and G. Fuchs.** 2006. Study of an alternate glyoxylate cycle for acetate assimilation by *Rhodobacter sphaeroides*. *Mol. Microbiol.* **61**:297-309.
4. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-10.
5. **Anthony, C.** 2011. How half a century of research was required to understand bacterial growth on C-1 and C-2 compounds; the story of the serine cycle and the ethylmalonyl-CoA pathway. *Science Prog.* **94**:109-137.
6. **Ashiuchi, M., and H. Misono.** 1999. Biochemical evidence that *Escherichia coli* hyi (orf b0508, gip) gene encodes hydroxypyruvate isomerase. *Biochim. Biophys. Acta* **1435**:153-159.
7. **ATSDR.** 1989. Toxicological profile for n-nitrosodimethylamine. Agency for Toxic Substances and Disease Registry.
8. **Barke, J., R. Seipke, S. Gruschow, D. Heavens, N. Drou, M. Bibb, R. Goss, D. Yu, and M. Hutchings.** 2010. A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant *Acromyrmex octospinosus*. *BMC Biol.* **8**:109.
9. **Bauchop, T., and S. R. Elsdén.** 1960. The growth of micro-organisms in relation to their energy supply. *J. Gen. Microbiol.* **23**:457-69.
10. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**:783-95.
11. **Benjamini, Y., and Y. Hochberg.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series. B Stat. Methodol.* **57**:289-300.
12. **Bennett, S.** 2004. Solexa Ltd. *Pharmacogenomics* **5**:433-438.
13. **Bernhardt, D., and H. Diekmann.** 1991. Degradation of dioxane, tetrahydrofuran and other cyclic ethers by an environmental *Rhodococcus* strain. *Appl. Microbiol. Biotech.* **36**:120-123.
14. **Bertoni, G., F. Bolognese, E. Galli, and P. Barbieri.** 1996. Cloning of the genes for and characterization of the early stages of toluene and *o*-xylene catabolism in *Pseudomonas stutzeri* OX1. *Appl. Environ. Microbiol.* **62**:3704.
15. **Bolstad, B. M., R. A. Irizarry, M. Åstrand, and T. P. Speed.** 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**:185-193.
16. **Boronat, A., E. Caballero, and J. Aguilar.** 1983. Experimental evolution of a metabolic pathway for ethylene glycol utilization by *Escherichia coli*. *J. Bacteriol.* **153**:134-139.
17. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248-254.
18. **Brown, K. L., and K. T. Hughes.** 1995. The role of anti-sigma factors in gene regulation. *Mol. Microbiol.* **16**:397.

19. **Burback, B. L., and J. J. Perry.** 1993. Biodegradation and biotransformation of groundwater pollutant mixtures by *Mycobacterium vaccae*. *Appl. Environ. Microbiol.* **59**:1025-1029.
20. **Cafaro, M. J., M. Poulsen, A. E. F. Little, S. L. Price, N. M. Gerardo, B. Wong, A. E. Stuart, B. Larget, P. Abbot, and C. R. Currie.** 2005. Specificity in the symbiotic association between fungus-growing ants and protective *Pseudonocardia* bacteria. *Proc. R. Soc. B. Biol. Sci.*
21. **California Department of Public Health.** 2011. A brief history of NDMA findings in drinking water. [Online.] Accessed February 3, 2012. (<http://www.cdph.ca.gov/certlic/drinkingwater/Pages/NDMAhistory.aspx>)
22. **California Department of Public Health.** 2011. NDMA and other nitrosamines. [Online.] Accessed February 3, 2012. (<http://www.cdph.ca.gov/certlic/drinkingwater/Pages/NDMA.aspx>)
23. **Chen, H.-H., S. Qin, J. Li, Y.-Q. Zhang, L.-H. Xu, C.-L. Jiang, C.-J. Kim, and W.-J. Li.** 2009. *Pseudonocardia endophytica* sp. nov., isolated from the pharmaceutical plant *Lobelia clavata*. *Int. J. Syst. Evol. Microbiol.* **59**:559-563.
24. **Colby, J., H. Dalton, and R. Whittenbury.** 1975. Improved assay for bacterial methane monooxygenase - some properties of enzyme from *Methylomonas methanica*. *Biochem. J.* **151**:459-462.
25. **Coleman, N. V., T. E. Mattes, J. M. Gossett, and J. C. Spain.** 2002. Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. *Appl. Environ. Microbiol.* **68**:6162-6171.
26. **Conway, T., and L. O. Ingram.** 1989. Similarity of *Escherichia coli* propanediol oxidoreductase (*fucO* product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *J. Bacteriol.* **171**:3754-3759.
27. **Cooper, R. A., and M. A. Skinner.** 1980. Catabolism of 3- and 4-hydroxyphenylacetate by the 3,4-dihydroxyphenylacetate pathway in *Escherichia coli*. *J. Bacteriol.* **143**:302-6.
28. **Cusa, E., N. Obradors, L. Baldoma, J. Badia, and J. Aguilar.** 1999. Genetic analysis of a chromosomal region containing genes required for assimilation of allantoin nitrogen and linked glyoxylate metabolism in *Escherichia coli*. *J. Bacteriol.* **181**:7479-7484.
29. **Dean-Raymond, D., and M. Alexander.** 1976. PLANT UPTAKE AND LEACHING OF DIMETHYLNITROSAMINE. *Nature* **262**:394-396.
30. **DeRosa, C. T., S. Wilbur, J. Holler, P. Richter, and Y. W. Stevens.** 1996. Health evaluation of 1,4-dioxane. *Toxicol. Ind. Health* **12**:1-43.
31. **Díaz, E.** 2004. Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *Int. Microbiol.* **7**:173-80.
32. **Draper, W. M., J. S. Dhoot, J. W. Remoy, and S. K. Perera.** 2000. Trace-level determination of 1,4-dioxane in water by isotopic dilution GC and GC-MS. *Analyst* **125**:1403-1408.
33. **Erb, T. J., I. A. Berg, V. Brecht, M. Müller, G. Fuchs, and B. E. Alber.** 2007. Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: The ethylmalonyl-CoA pathway. *Proc. Nat. Acad. Sci.* **104**:10631-10636.
34. **Erb, T. J., V. Brecht, G. Fuchs, M. Mueller, and B. E. Alber.** 2009. Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proc. Nat. Acad. Sci.* **106**:8871-8876.
35. **Erb, T. J., L. Frerichs-Revermann, G. Fuchs, and B. E. Alber.** 2010. The apparent malate synthase activity of *Rhodobacter sphaeroides* is due to two paralogous enzymes, (3S)-methyl-Coenzyme A (CoA)/beta-methylmalyl-CoA lyase and (3S)-malyl-CoA thioesterase. *J. Bacteriol.* **192**:1249-1258.

36. **Erb, T. J., G. Fuchs, and B. E. Alber.** 2009. (2S)-Methylsuccinyl-CoA dehydrogenase closes the ethylmalonyl-CoA pathway for acetyl-CoA assimilation. *Mol. Microbiol.* **73**:992-1008.
37. **Erb, T. J., J. Retey, G. Fuchs, and B. E. Alber.** 2008. Ethylmalonyl-CoA mutase from *Rhodobacter sphaeroides* defines a new subclade of Coenzyme B-12-dependent acyl-CoA mutases. *J. Biol. Chem.* **283**:32283-32293.
38. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Gen. Res.* **8**:175-185.
39. **Fournier, D., J. Hawari, S. H. Streger, K. McClay, and P. B. Hatzinger.** 2006. Biotransformation of N-nitrosodimethylamine by *Pseudomonas mendocina* KR1. *Appl. Environ. Microbiol.* **72**:6693-6698.
40. **Galán, B., E. Díaz, M. A. Prieto, and J. L. García.** 2000. Functional analysis of the small component of the 4-hydroxyphenylacetate 3-monooxygenase of *Escherichia coli* W: a prototype of a new Flavin:NAD(P)H reductase subfamily. *J. Bacteriol.* **182**:627-36.
41. **Gautier, L., L. Cope, B. M. Bokstad, and R. A. Irizarry.** 2004. affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**:307-315.
42. **Griffiths-Jones, S., A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy.** 2003. Rfam: an RNA family database. *Nucleic Acids Res.* **31**:439-41.
43. **Gu, Q., H. Luo, W. Zheng, Z. Liu, and Y. Huang.** 2006. *Pseudonocardia oroxyli* sp. nov., a novel actinomycete isolated from surface-sterilized *Oroxylum indicum* root. *Int. J. Syst. Evol. Microbiol.* **56**:2193-2197.
44. **Gunnison, D., M. E. Zappi, C. Teeter, J. C. Pennington, and R. Bajpai.** 2000. Attenuation mechanisms of N-nitrosodimethylamine at an operating intercept and treat groundwater remediation system. *J. Hazard Mater.* **73**:179-197.
45. **Gürtler, V., B. C. Mayall, and R. Seviour.** 2004. Can whole genome analysis refine the taxonomy of the genus *Rhodococcus*? *FEMS Micro. Rev.* **28**:377-403.
46. **Hamamura, N., C. Page, T. Long, L. Semprini, and D. J. Arp.** 1997. Chloroform cometabolism by butane-grown CF8, *Pseudomonas butanovora*, and *Mycobacterium vaccae* JOB5 and methane-grown *Methylosinus trichosporium* OB3b. *Appl. Environ. Microbiol.* **63**:3607-3613.
47. **Hamamura, N., C. M. Yeager, and D. J. Arp.** 2001. Two distinct monooxygenases for alkane oxidation in *Nocardioide*s sp. strain CF8. *Appl. Environ. Microbiol.* **67**:4992-4998.
48. **Han, C., and P. Chain.** 2006. Finishing repeat regions automatically with Dupfinisher. Presented at the Proc. 2006 Int. Conf. Bioinformatics Comp.Biol., 2006.
49. **Hansen, R. W., and J. A. Hayashi.** 1962. Glycolate metabolism in *Escherichia coli*. *J. Bacteriol.* **83**:679-687.
50. **Harayama, S., M. Kok, and E. L. Neidle.** 1992. Functional and evolutionary relationships among diverse oxygenases. *Ann. Rev. Microbiol.* **46**:565-601.
51. **Helmann, J. D.** 2002. The extracytoplasmic function (ECF) sigma factors. *Adv. Microbial Phys.* **46**:47-110.
52. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. *Ann. Rev. Biochem.* **57**:839-72.
53. **Hidalgo, E., Y. M. Chen, E. C. Lin, and J. Aguilar.** 1991. Molecular cloning and DNA sequencing of the *Escherichia coli* K-12 ald gene encoding aldehyde dehydrogenase. *J. Bacteriol.* **173**:6118-6123.
54. **Hill, R. R., G. E. Jeffs, and D. R. Roberts.** 1997. Photocatalytic degradation of 1,4-dioxane in aqueous solution. *J. Photochem. Photobiol. A Chem.* **108**:55-58.
55. **Hyatt, D., G.-L. Chen, P. LoCascio, M. Land, F. Larimer, and L. Hauser.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119.

56. **IARC.** 1999. 1,4-Dioxane: Summary of data reported and evaluation. [Online.] Accessed February 2, 2012. (<http://www.inchem.org/documents/iarc/vol71/019-dioxane.html>)
57. **Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed.** 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**:15.
58. **Jackson, R. E., and V. Dwarakanath.** 1999. Chlorinated degreasing solvents: Physical-chemical properties affecting aquifer contamination and remediation. *Ground Water Monit. Remediat.* **19**:102-110.
59. **Johns, M. M., W. E. Marshall, and C. A. Toles.** 1998. Agricultural by-products as granular activated carbons for adsorbing dissolved metals and organics. *J. Chem. Technol. Biotech.* **71**:131-140.
60. **Johnson, D. R., E. L. Brodie, A. E. Hubbard, G. L. Andersen, S. H. Zinder, and L. Alvarez-Cohen.** 2008. Temporal transcriptomic microarray analysis of "*Dehalococcoides ethenogenes*" strain 195 during the transition into stationary phase. *Appl. Environ. Microbiol.* **74**:2864-2872.
61. **Kalscheuer, R., M. Arenskötter, and A. Steinbüchel.** 1999. Establishment of a gene transfer system for *Rhodococcus opacus* PD630 based on electroporation and its application for recombinant biosynthesis of poly(3-hydroxyalkanoic acids). *Appl. Environ. Microbiol.* **52**:508-515.
62. **Kampfer, P., and R. M. Kroppenstedt.** 2004. *Pseudonocardia benzenivorans* sp. nov. *Int. J. Syst. Evol. Microbiol.* **54**:749-51.
63. **Kaplan, D. L., and A. M. Kaplan.** 1985. Biodegradation of n-nitrosodimethylamine in aqueous and soil systems. *Appl. Environ. Microbiol.* **50**:1077-1086.
64. **Karzai, A. W., E. D. Roche, and R. T. Sauer.** 2000. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat. Struct. Biol.* **7**:449-455.
65. **Keiler, K. C., P. R. H. Waller, and R. T. Sauer.** 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**:990-993.
66. **Kim, Y.-M., J.-R. Jeon, K. Murugesan, E.-J. Kim, and Y.-S. Chang.** 2009. Biodegradation of 1,4-dioxane and transformation of related cyclic compounds by a newly isolated *Mycobacterium* sp. PH-06. *Biodegradation* **20**:511-519.
67. **Kochi, H., and G. Kikuchi.** 1969. Reactions of glycine synthesis and glycine cleavage catalyzed by extracts of *Arthrobacter globiformis* grown on glycine. *Arch. Biochem. Biophys.* **132**:359-369.
68. **Kohlweyer, U., B. Thiemer, T. Schrader, and J. R. Andreessen.** 2000. Tetrahydrofuran degradation by a newly isolated culture of *Pseudonocardia* sp. strain K1. *FEMS Micro. Lett.* **186**:301-6.
69. **Kornberg, H. L.** 1966. Anaplerotic sequences and their role in metabolism [Bacteria, Fungi, Protozoa]. *Essays Biochem.* **2**:1-31.
70. **Kornberg, H. L., and S. R. Elsdén.** 1961. The metabolism of 2-carbon compounds by microorganisms, p. 401-470, *Advances in Enzymology and Related Areas of Molecular Biology*. John Wiley & Sons, Inc.
71. **Kornberg, H. L., and H. A. Krebs.** 1957. Synthesis of cell constituents from C₂-units by a modified tricarboxylic acid cycle. *Nature* **179**.
72. **Kotani, T., T. Yamamoto, H. Yurimoto, Y. Sakai, and N. Kato.** 2003. Propane monooxygenase and NAD⁺-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. strain TY-5. *J. Bacteriol.* **185**:7120-7128.
73. **Krakov, G., S. S. Barkulis, and J. A. Hayashi.** 1961. Glyoxylic acid carboligase: an enzyme present in glycolate-grown *Escherichia coli* J. *Bacteriol.* **81**:509-518.
74. **Kroeger-Koepke, M. B., A. W. Andrews, R. J. Kupper, S. R. Koepke, and C. J. Michejda.** 1981. Mutagenicity and rat-liver S9 demethylation kinetics of n-nitrosomethylaniline and its ring-substituted derivatives. *Mut. Res.* **89**:255-267.

75. **Kroeger-Koepke, M. B., S. R. Koepke, G. A. McClusky, P. N. Magee, and C. J. Michejda.** 1981. Alpha-hydroxylation pathway in the *in vitro* metabolism of carcinogenic nitrosamines - n-nitrosodimethylamine and n-nitroso-n-methylaniline. *Proc. Nat. Acad. Sci.* **78**:6489-6493.
76. **Krogh, A., B. r. Larsson, G. von Heijne, and E. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**:567-580.
77. **Kumar, R., R. Saini, R. Kapoor, T. O. Siddiqi, and A. Kumar.** 2011. CO₂ utilizing microbes - A comprehensive review. *Biotechnol. Adv.* **29**:949-960.
78. **Lagesen, K., P. Hallin, E. A. Rødland, H.-H. Staerfeldt, T. r. Rognes, and D. W. Ussery.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**:3100-3108.
79. **Langille, M. G. I., and F. S. L. Brinkman.** 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25**:664-665.
80. **Larkin, M. J., R. De Mot, L. A. Kulakov, and I. Nagy.** 1998. Applied aspects of *Rhodococcus* genetics. *Antonie van Leeuwenhoek* **74**:133-153.
81. **Lee, C., I. Kim, J. Lee, K.-L. Lee, B. Min, and C. Park.** 2010. Transcriptional activation of the aldehyde reductase YqhD by YqhC and its implication in glyoxal metabolism of *Escherichia coli* K-12. *J. Bacteriol.* **192**:4205-4214.
82. **Lee, M.-Y., J. Myeong, H.-J. Park, K. Han, and E.-S. Kim.** 2006. Isolation and partial characterization of a cryptic polyene gene cluster in *Pseudonocardia autotrophica*. *J. Ind. Microbiol. Biotechnol.* **33**:84-87.
83. **Lee, S. B., S. E. Strand, H. D. Stensel, and R. P. Herwig.** 2004. *Pseudonocardia chloroethenivorans* sp. nov., a chloroethene-degrading actinomycete. *Int. J. Syst. Evol. Microbiol.* **54**:131-139.
84. **Li, F., C. H. Hagemeyer, H. Seedorf, G. Gottschalk, and R. K. Thauer.** 2007. Re-citrate synthase from *Clostridium kluyveri* is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to Si-citrate synthase. *J. Bacteriol.* **189**:4299-4304.
85. **Liu, Z.** 1996. Hetero-stagger cloning: efficient and rapid cloning of PCR products. *Nucleic Acids Res.* **24**:2458-2459.
86. **Lopes Ferreira, N., H. Mathis, D. Labbe, F. Monot, C. W. Greer, and F. Fayolle-Guichard.** 2007. n-Alkane assimilation and tert-butyl alcohol (TBA) oxidation capacity in *Mycobacterium austroafricanum* strains. *Appl. Microbiol. Biotech.* **75**:909-919.
87. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955-64.
88. **Ma, Y. F., Y. Zhang, J. Y. Zhang, D. W. Chen, C. Y. Jiang, S. J. Liu, Y. Zhu, H. Zheng, S. Y. Wang, and G. P. Zhao.** 2009. The complete genome of *Comamonas testosteroni* reveals its genetic adaptations to changing environments. *Appl. Environ. Microbiol.* **75**:6812-6819.
89. **Mahendra, S., and L. Alvarez-Cohen.** 2006. Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. *Environ. Sci. Technol.* **40**:5435-5442.
90. **Mahendra, S., and L. Alvarez-Cohen.** 2005. *Pseudonocardia dioxanivorans* sp. nov., a novel actinomycete that grows on 1,4-dioxane. *Int. J. Syst. Evol. Microbiol.* **55**:593-598.
91. **Mahendra, S., C. J. Petzold, E. E. Baidoo, J. D. Keasling, and L. Alvarez-Cohen.** 2007. Identification of the intermediates of *in vivo* oxidation of 1,4-dioxane by monooxygenase-containing bacteria. *Environ. Sci. Technol.* **41**:7330-7336.
92. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M.**

- L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376-380.
93. Masuda, H. 2009. Identification and characterization of monooxygenase enzymes involved in 1,4-dioxane degradation in *Pseudonocardia* sp. strain ENV478, *Mycobacterium* sp. strain ENV421, and *Nocardia* sp. strain ENV425. Rutgers University, New Brunswick, NJ, .
94. McClay, K., B. G. Fox, and R. J. Steffan. 2000. Toluene monooxygenase-catalyzed epoxidation of alkenes. *Appl. Environ. Microbiol.* **66**:1877-1882.
95. McFadden, B. A. 1969. Isocitrate lyase, p. 163-170. In M. L. John (ed.), *Meth. Enzym.*, vol. Volume 13. Academic Press.
96. McLeod, M. P., R. L. Warren, W. W. L. Hsiao, N. Araki, M. Myhre, C. Fernandes, D. Miyazawa, W. Wong, A. L. Lillquist, D. Wang, M. Dosanjh, H. Hara, A. Petrescu, R. D. Morin, G. Yang, J. M. Stott, J. E. Schein, H. Shin, D. Smailus, A. S. Siddiqui, M. A. Marra, S. J. M. Jones, R. Holt, F. S. L. Brinkman, K. Miyauchi, M. Fukuda, J. E. Davies, W. W. Mohn, and L. D. Eltis. 2006. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Nat. Acad. Sci.* **103**:15582-15587.
97. Mirvish, S. S. 1975. Blocking formation of n-nitroso compounds with ascorbic acid *in vitro* and *in vivo*. *Ann. N.Y. Acad. Sci.* **258**:175-180.
98. Mohr, T. K. G. 2001. 1,4-Dioxane and other solvent stabilizers. White paper. Santa Clara Valley Water District.
99. Molina, I., M.-T. Pellicer, J. Badia, J. Aguilar, and L. Baldoma. 1994. Molecular characterization of *Escherichia coli* malate synthase G. *Eur. J. Biochem.* **224**:541-548.
100. Najm, I., and R. R. Trussell. 2001. NDMA formation in water and wastewater. *J. Am. Water Works Assoc.* **93**:92-99.
101. Nakamiya, K., S. Hashimoto, H. Ito, J. S. Edmonds, and M. Morita. 2005. Degradation of 1,4-dioxane and cyclic ethers by an isolated fungus. *Appl. Environ. Microbiol.* **71**:1254-1258.
102. Nakashima, N., and T. Tamura. 2004. Isolation and characterization of a rolling-circle-type plasmid from *Rhodococcus erythropolis* and application of the plasmid to multiple-recombinant-protein expression. *Appl. Environ. Microbiol.* **70**:5557-5568.
103. Nebert, D. W., D. R. Nelson, M. J. Coon, R. W. Estabrook, R. Feyereisen, Y. Fujikuriyama, F. J. Gonzalez, F. P. Guengerich, I. C. Gunsalus, E. F. Johnson, J. C. Loper, R. Sato, M. R. Waterman, and D. J. Waxman. 1991. The P450 superfamily - update on new sequences, gene-mapping and recommended nomenclature. *DNA Cell Biol.* **10**:1-14.
104. O'Keefe, D. P., and P. A. Harder. 1991. Occurrence and biological function of cytochrome-P450 monooxygenases in the actinomycetes. *Mol. Microbiol.* **5**:2099-2105.
105. Okubo, Y., S. Yang, L. Chistoserdova, and M. E. Lidstrom. 2010. Alternative route for glyoxylate consumption during growth on two-carbon compounds by *Methylobacterium extorquens* AM1. *J. Bacteriol.* **192**:1813-1823.
106. Oldenhuis, R., R. Vink, D. B. Janssen, and B. Witholt. 1989. Degradation of chlorinated aliphatic hydrocarbons by *Methylosinus trichosporium* Ob3B expressing soluble methane monooxygenase. *Appl. Environ. Microbiol.* **55**:2819-2826.
107. Oliver, J. E. 1979. Volatilization of some herbicide-related nitrosamines from soils, p. 596-601, *J. Environ. Qual.*, vol. 8.

108. **Oliynyk, M., M. Samborskyy, J. B. Lester, T. Mironenko, N. Scott, S. Dickens, S. F. Haydock, and P. F. Leadlay.** 2007. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotech.* **25**:447-453.
109. **Orange Country Water District.** 2002. Orange Country Water District takes a proactive stance on contaminants of concern. [Online.] Accessed February 3, 2012. (http://www.ocwd.com/_html/_pr/_pr02/pr02_0129_dioxane.htm)
110. **Ornston, L. N., and M. K. Ornston.** 1969. Regulation of glyoxylate metabolism in *Escherichia coli* K-12. *J. Bacteriol.* **98**:1098-1108.
111. **Pan, Y., X. Yang, J. Li, R. Zhang, Y. Hu, Y. Zhou, J. Wang, and B. Zhu.** 2011. The genome sequence of spinosyns-producing bacterium *Saccharopolyspora spinosa* NRRL 18395. *J. Bacteriol.*
112. **Parales, R. E., J. E. Adamus, N. White, and H. D. May.** 1994. Degradation of 1,4-dioxane by an actinomycete in pure culture. *Appl. Environ. Microbiol.* **60**:4527-30.
113. **Pati, A., N. N. Ivanova, N. Mikhailova, G. Ovchinnikova, S. D. Hooper, A. Lykidis, and N. C. Kyrpides.** 2010. GenePRIMP: A gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* **7**:455-457.
114. **Pati, A., J. Sikorski, M. Nolan, A. Lapidus, A. Copeland, T. Glavina Del Rio, S. Lucas, F. Chen, H. Tice, S. Pitluck, J. F. Cheng, O. Chertkov, T. Brettin, C. Han, J. C. Detter, C. Kuske, D. Bruce, L. Goodwin, P. Chain, P. D'Haeseleer, A. Chen, K. Palaniappan, N. Ivanova, K. Mavromatis, N. Mikhailova, M. Rohde, B. J. Tindall, M. Goker, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, and H. P. Klenk.** 2009. Complete genome sequence of *Saccharomonospora viridis* type strain (P101). *Stand. Genomic Sci.* **1**:141-149.
115. **Patt, T. E., and H. M. Abebe.** March 21, 1995 1995. Microbial degradation of chemical pollutants patent US 5399495.
116. **Pellicer, M. T., J. Badia, J. Aguilar, and L. Baldoma.** 1996. *glc* locus of *Escherichia coli*: characterization of genes encoding the subunits of glycolate oxidase and the *glc* regulator protein. *J. Bacteriol.* **178**:2051-2059.
117. **Pellicer, M. T., C. Fernandez, J. Badía, J. Aguilar, E. C. C. Lin, and L. Baldomà** 1999. Cross-induction of *glc* and *ace* operons of *Escherichia coli* attributable to pathway intersection. *J. Biol. Chem.* **274**:1745-1752.
118. **Peyraud, R., P. Kiefer, P. Christen, S. Massou, J.-C. Portais, and J. A. Vorholt.** 2009. Demonstration of the ethylmalonyl-CoA pathway by using (13)C metabolomics. *Proc. Nat. Acad. Sci.* **106**:4846-4851.
119. **Poulsen, M., M. J. Cafaro, D. P. Erhardt, A. E. F. Little, N. M. Gerardo, B. Tebbets, B. S. Klein, and C. R. Currie.** 2005. Variation in *Pseudonocardia* antibiotic defence helps govern parasite-induced morbidity in *Acromyrmex* leaf-cutting ants. *Environ. Microbiol. Rep.* **2**:534-540.
120. **Prabahar, V., S. Dube, G. S. N. Reddy, and S. Shivaji.** 2004. *Pseudonocardia antarctica* sp. nov., an *Actinomycetes* from McMurdo Dry Valleys, Antarctica. *Syst. Appl. Microbiol.* **27**:66-71.
121. **Pridmore, R. D.** 1987. New and versatile cloning vectors with kanamycin-resistance marker. *Gene* **56**:309-312.
122. **Prieto, M. A., and J. L. Garcia.** 1994. Molecular characterization of 4-hydroxyphenylacetate 3-hydroxylase of *Escherichia coli*. A two-protein component enzyme. *J. Biol. Chem.* **269**:22823-9.
123. **Prior, S. D., and H. Dalton.** 1985. Acetylene as a suicide substrate and active-site probe for methane monooxygenase from *Methylococcus capsulatus* (Bath). *FEMS Microbiol. Lett.* **29**:105-109.

124. **Qin, S., W.-Y. Zhu, J.-H. Jiang, H.-P. Klenk, J. Li, G.-Z. Zhao, L.-H. Xu, and W.-J. Li.** 2010. *Pseudonocardia tropica* sp. nov., an endophytic actinomycete isolated from the stem of *Maytenus austroyunnanensis*. Int. J. Syst. Evol. Microbiol. **60**:2524-2528.
125. **Raj, C. B. C., N. Ramkumar, A. H. J. Siraj, and S. Chidambaram.** 1997. Biodegradation of acetic, benzoic, isophthalic, toluic and terephthalic acids using a mixed culture: Effluents of PTA production. Proc. Safety Environ. Prot. **75**:245-256.
126. **Ribbe, M., D. Gadkari, and O. Meyer.** 1997. N₂ fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-superoxide oxidoreductase that couple N₂ reduction to the oxidation of superoxide produced from O₂ by a molybdenum-CO dehydrogenase. J. Biol. Chem. **272**:26627-33.
127. **Roy, D., G. Anagnostu, and P. Chaphalkar.** 1994. Biodegradation of dioxane and diglyme in industrial waste. J. Environ. Sci. Health. A Tox. Hazard Subst. Environ. Eng. **29**:129.
128. **Russell, J. B., and G. M. Cook.** 1995. Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol. Rev. **59**:48-62.
129. **Sagers, R. D., and I. C. Gunsalus.** 1961. Intermediary metabolism of *Diplococcus glyciniphilus* I.: Glycine cleavage and one-carbon interconversions. J. Bacteriol. **81**:541-549.
130. **Sakai, A., K. Katayama, T. Katsuragi, and Y. Tani.** 2001. Glycolaldehyde-forming route in *Bacillus subtilis* in relation to vitamin B6 biosynthesis. J. Biosci. Bioeng. **91**:147-152.
131. **Sales, C. M.** 2012. Functional genomics of bacterial degradation of the emerging contaminants 1,4-dioxane and N-nitrosodimethylamine (NDMA). Ph.D. thesis. University of California, Berkeley, Berkeley, CA.
132. **Sales, C. M., S. Mahendra, A. Grostern, R. E. Parales, L. A. Goodwin, T. Woyke, M. Nolan, A. Lapidus, O. Chertkov, G. Ovchinnikova, A. Sczyrba, and L. Alvarez-Cohen.** 2011. Genome sequence of the 1,4-dioxane-degrading *Pseudonocardia dioxanivorans* strain CB1190. J. Bacteriol. **193**:4549-4550.
133. **Sallach, H. J.** 1956. Formation of serine from hydroxypyruvate and L-alanine. J. Biol. Chem. **223**:1101-1108.
134. **Sedlmeier, R., and J. Altenbuchner.** 1992. Cloning and DNA sequence analysis of the mercury resistance genes of *Streptomyces lividans*. Mol. Gen. Genet. **236**:76-85.
135. **Shao, Z., J. Gao, X. Ding, J. Wang, J. Chiao, and G. Zhao.** 2011. Identification and functional analysis of a nitrate assimilation operon *nasACKBDEF* from *Amycolatopsis mediterranei* U32. Arch. Microbiol. **193**:463-477.
136. **Sharp, J. O.** 2006. Aerobic bacterial degradation of the water pollutant N-nitrosodimethylamine (NDMA). Ph.D. thesis. University of California, Berkeley, Berkeley, CA.
137. **Sharp, J. O., C. M. Sales, J. C. LeBlanc, J. Liu, T. K. Wood, L. D. Eltis, W. W. Mohn, and L. Alvarez-Cohen.** 2007. An inducible propane monooxygenase is responsible for N-Nitrosodimethylamine degradation by *Rhodococcus* sp. strain RHA1. Appl. Environ. Microbiol. **73**:6930-6938.
138. **Sharp, J. O., T. K. Wood, and L. Alvarez-Cohen.** 2005. Aerobic biodegradation of n-nitrosodimethylamine (NDMA) by axenic bacterial strains. Biotechnol. Bioeng. **89**:608-618.
139. **Shen, W., H. Chen, and S. Pan.** 2008. Anaerobic biodegradation of 1,4-dioxane by sludge enriched with iron-reducing microorganisms. Biores. Technol. **99**:2483-2487.
140. **Singh, R., J. Lemire, R. J. Mailloux, D. Chénier, R. Hamel, and V. D. Appanna.** 2009. An ATP and oxalate generating variant tricarboxylic acid cycle counters aluminum toxicity in *Pseudomonas fluorescens*. PLoS ONE **4**:e7344.

141. **Skinner, K., L. Cuiffetti, and M. Hyman.** 2009. Metabolism and cometabolism of cyclic ethers by a filamentous fungus, a *Graphium* sp. Appl. Environ. Microbiol. **75**:5514-5522.
142. **Smith, C. A., K. T. O'Reilly, and M. R. Hyman.** 2003. Characterization of the initial reactions during the cometabolic oxidation of methyl tert-butyl ether by propane-grown *Mycobacterium vaccae* JOB5. Appl. Environ. Microbiol. **69**:796-804.
143. **Smyth, G.** 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. **3**.
144. **Stefan, M. I., and J. R. Bolton.** 1998. Mechanism of the degradation of 1,4-dioxane in dilute aqueous solution using the UV hydrogen peroxide process. Environ. Sci. Technol. **32**:1588-1595.
145. **Stiborova, M., H. H. Schmeiser, and E. Frei.** 2000. Oxidation of xenobiotics by plant microsomes, a reconstituted cytochrome P450 system and peroxidase: a comparative study. Phytochemistry **54**:353-362.
146. **Stickney, J. A., S. L. Sager, J. R. Clarkson, L. A. Smith, B. J. Locey, M. J. Bock, R. Hartung, and S. F. Olp.** 2003. An updated evaluation of the carcinogenic potential of 1,4-dioxane. Regul. Toxicol. Pharmacol. **38**:183-195.
147. **Stirling, D. I., and H. Dalton.** 1979. Properties of the methane mono-oxygenase from extracts of *Methylosinus trichosporium* Ob3B and evidence for its similarity to the enzyme from *Methylococcus capsulatus* (bath) Eur. J. Biochem. **96**:205-212.
148. **Sun, B., K. Ko, and J. A. Ramsay.** 2011. Biodegradation of 1,4-dioxane by a *Flavobacterium*. Biodegradation **22**:651-659.
149. **Tang, Y. J., S. Yi, W. Q. Zhuang, S. H. Zinder, J. D. Keasling, and L. Alvarez-Cohen.** 2009. Investigation of carbon metabolism in "Dehalococcoides ethenogenes" strain 195 by use of isotopomer and transcriptomic analyses. J. Bacteriol. **191**:5224-5231.
150. **Tanner, A., and S. Bornemann.** 2000. *Bacillus subtilis* YvrK is an acid-induced oxalate decarboxylase. J. Bacteriol. **182**:5271-5273.
151. **Tate, R. L., and M. Alexander.** 1975. Stability of nitrosamines in samples of lake water, soil and sewage. J. Natl. Cancer Inst. **54**:327-330.
152. **te Poele, E. M., M. Samborsky, M. Oliynyk, P. F. Leadlay, H. Bolhuis, and L. Dijkhuizen.** 2008. Actinomycete integrative and conjugative pMEA-like elements of *Amiclatopsis* and *Saccharopolyspora* decoded. Plasmid **59**:202-16.
153. **Thierner, B., J. R. Andreesen, and T. Schröder.** 2001. The NADH-dependent reductase of a putative multicomponent tetrahydrofuran mono-oxygenase contains a covalently bound FAD. Eur. J. Biochem. **268**:3774-3782.
154. **Thierner, B., J. R. Andreesen, and T. Schröder.** 2003. Cloning and characterization of a gene cluster involved in tetrahydrofuran degradation in *Pseudonocardia* sp. strain K1. Arch. Microbiol. **179**:266-277.
155. **Thornalley, P. J.** 1998. Glutathione-dependent detoxification of alpha-oxoaldehydes by the glyoxalase system: involvement in disease mechanisms and antiproliferative activity of glyoxalase I inhibitors. Chem. Biol. Interact. **111-112**:137-51.
156. **Tran, C. V.** 2004. TCDB : a membrane transport protein classification database. Dissertation/Thesis.
157. **Tu, Y. Y., and C. S. Yang.** 1985. Demethylation and denitrosation of nitrosamines by cytochrome P-450 isozymes. Arch. Biochem. Biophys. **242**:32-40.
158. **Ulrich, L. E., E. V. Koonin, and I. B. Zhulin.** 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. **13**:52-6.
159. **Ulrich, L. E., and I. B. Zhulin.** 2010. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res. **38**:401-7.
160. **USEPA.** 2001. Record of decision for the Western Groundwater Operable Unit OU-3, Aerojet Sacramento site. United States Environmental Protection Agency.

161. **Vainberg, S., K. McClay, H. Masuda, D. Root, C. Condee, G. J. Zylstra, and R. J. Steffan.** 2006. Biodegradation of ether pollutants by *Pseudonocardia* sp. strain ENV478. *Appl. Environ. Microbiol.* **72**:5218-5224.
162. **van der Hoeven, J. S., and C. W. A. van den Kieboom.** 1990. Oxygen-dependent lactate utilization by *Actinomyces viscosus* and *Actinomyces naeslundii*. *Oral Microbiol. Immunol.* **5**:223-225.
163. **Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**:research0034-research0034.11
164. **Verma, M., J. Kaur, M. Kumar, K. Kumari, A. Saxena, S. Anand, A. Nigam, V. Ravi, S. Raghuvanshi, P. Khurana, A. K. Tyagi, J. P. Khurana, and R. Lal.** 2011. Whole genome sequence of the rifamycin B-producing strain *Amycolatopsis mediterranei* S699. *J. Bacteriol.* **193**:5562-5563.
165. **Wackett, L. P., G. A. Brusseau, S. R. Householder, and R. S. Hanson.** 1989. Survey of microbial oxygenases: trichloroethylene degradation by propane-oxidizing bacteria. *Appl. Environ. Microbiol.* **55**:2960-4.
166. **Washtenaw County.** 2011. CARD (1,4-Dioxane). [Online.] Accessed February 3, 2012. (http://www.ewashtenaw.org/government/departments/environmental_health/card)
167. **White, G. F., N. J. Russell, and E. C. Tidswell.** 1996. Bacterial scission of ether bonds. *Microbiol. Rev.* **60**:216-232.
168. **Woo, Y.-t., J. C. Arcos, M. F. Argus, G. W. Griffin, and K. Nishiyama.** 1977. Metabolism of dioxane: identification of *p*-dioxane-2-one as the major urinary metabolite. *Biochem. Pharmacol.* **26**:1535-1538.
169. **Woo, Y. T., J. C. Arcos, M. F. Argus, G. W. Griffin, and K. Nishiyama.** 1977. Structural identification of *p*-dioxane-2-one as the major urinary metabolite of *p*-dioxane. *Naunyn Schmiedebergs Arch. Pharmacol.* **299**:283-7.
170. **Woo, Y. T., M. F. Argus, and J. C. Arcos.** 1977. Metabolism in vivo of dioxane: effect of inducers and inhibitors of hepatic mixed-function oxidases. *Biochem. Pharmacol.* **26**:1539-42.
171. **Wood, H. G., and M. F. Utter.** 1965. The role of CO₂ fixation in metabolism. *Essays Biochem.* **1**:1-27.
172. **Yamashita, M., A. Tani, and F. Kawai.** 2004. Cloning and expression of an ether-bond-cleaving enzyme involved in the metabolism of polyethylene glycol. *J. Biosci. Bioeng.* **98**:313-315.
173. **Yao, Y., Z. Lv, H. Min, Z. Lv, and H. Jiao.** 2009. Isolation, identification and characterization of a novel *Rhodococcus* sp strain in biodegradation of tetrahydrofuran and its medium optimization using sequential statistics-based experimental designs. *Biores. Technol.* **100**:2762-2769.
174. **Yen, K. M., and M. R. Karl.** 1992. Identification of a new gene, *tmoF*, in the *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. *J. Bacteriol.* **174**:7253-61.
175. **Yen, K. M., M. R. Karl, L. M. Blatt, M. J. Simon, R. B. Winter, P. R. Fausset, H. S. Lu, A. A. Harcourt, and K. K. Chen.** 1991. Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. *J. Bacteriol.* **173**:5315-27.
176. **Yoshinari, T., and D. Shafer.** 1990. Degradation of dimethyl nitrosamine by *Methylosinus trichosporium* OB3b. *Can. J. Microbiol.* **36**:834-838.
177. **Young, J. D., W. H. Braun, and P. J. Gehring.** 1978. Dose-dependent fate of 1,4-dioxane in rats. *J. Toxicol. Environ. Health* **4**:709-726.

178. **Young, J. D., W. H. Braun, P. J. Gehring, B. S. Horvath, and R. L. Daniel.** 1976. 1,4-Dioxane and β -hydroxyethoxyacetic acid excretion in urine of humans exposed to dioxane vapors. *Toxicol. Appl. Pharmacol.* **38**:643-646.
179. **Zarzycki, J., A. Schlichting, N. Strychalsky, M. Mueller, B. E. Alber, and G. Fuchs.** 2008. Mesoconyl-coenzyme A hydratase, a new enzyme of two central carbon metabolic pathways in bacteria. *J. Bacteriol.* **190**:1366-1374.
180. **Zenker, M. J., R. C. Borden, and M. A. Barlaz.** 2000. Mineralization of 1,4-dioxane in the presence of a structural analog. *Biodegradation* **11**:239-246.
181. **Zenker, M. J., R. C. Borden, and M. A. Barlaz.** 2003. Occurrence and treatment of 1,4-dioxane in aqueous environments. *Environ. Eng. Sci.* **20**:423-432.
182. **Zerbino, D. R., and E. Birney.** 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**:821-829.
183. **Zhao, W., Y. Zhong, H. Yuan, J. Wang, H. Zheng, Y. Wang, X. Cen, F. Xu, J. Bai, X. Han, G. Lu, Y. Zhu, Z. Shao, H. Yan, C. Li, N. Peng, Z. Zhang, Y. Zhang, W. Lin, Y. Fan, Z. Qin, Y. Hu, B. Zhu, S. Wang, X. Ding, and G.-P. Zhao.** 2010. Complete genome sequence of the rifamycin SV-producing *Amycolatopsis mediterranei* U32 revealed its genetic characteristics in phylogeny and metabolism. *Cell Res.* **20**:1096-1108.
184. **Zhuang, W.-Q., S. Yi, X. Feng, S. H. Zinder, Y. J. Tang, and L. Alvarez-Cohen.** 2011. Selective utilization of exogenous amino acids by *Dehalococcoides ethenogenes* strain 195 and its effects on growth and dechlorination activity. *Appl. Environ. Microbiol.* **77**:7797-7803.

Appendix 1. List of scientific publication

Articles in peer-reviewed journals

- Grostern, A., C.M. Sales, W.-Q. Zhuang, O. Erbilgin, and L. Alvarez-Cohen. Glyoxylate metabolism is a key feature of the metabolic degradation of 1,4-dioxane by *Pseudonocardia dioxanivorans* strain CB1190. Accepted for publication in Applied and Environmental Microbiology, February 6, 2012.
- Mahendra, S., Grostern and L. Alvarez-Cohen. 2012. The impact of common co-contaminants and inducing substrates on degradation and inhibition kinetics of 1,4-dioxane. *Environ. Sci. Technol.*, submitted.
- Sales, C. M., S. Mahendra, A. Grostern, R. E. Parales, L. A. Goodwin, T. Woyke, M. Nolan, A. Lapidus, O. Chertkov, G. Ovchinnikova, A. Sczyrba, and L. Alvarez-Cohen. 2011. Genome sequence of the 1,4-dioxane-degrading *Pseudonocardia dioxanivorans* strain CB1190. *J. Bacteriol.* 193:4549-4550.
- Sharp, J. O., C. M. Sales, and L. Alvarez-Cohen. 2010. Functional characterization of propane-enhanced N-nitrosodimethylamine degradation by two actinomycetales. *Biotechnol. Bioeng.* 107:924-932.
- Sharp, J. O., C. M. Sales, J. C. LeBlanc, J. Liu, T. K. Wood, L. D. Eltis, W. W. Mohn, and L. Alvarez-Cohen. 2007. An inducible propane monooxygenase is responsible for N-Nitrosodimethylamine degradation by *Rhodococcus sp.* strain RHA1. *Appl. Environ. Microbiol.* 73:6930-6938.
- Mahendra, S., C. J. Petzold, E. E. Baidoo, J. D. Keasling, and L. Alvarez-Cohen. 2007. Identification of the intermediates of in vivo oxidation of 1,4-dioxane by monooxygenase-containing bacteria. *Environ. Sci. Technol.* 41:7330-7336.
- Mahendra, S., and L. Alvarez-Cohen. 2006. Kinetics of 1,4-dioxane biodegradation by monooxygenase-expressing bacteria. *Environ. Sci. Technol.* 40:5435-5442.

Conference abstracts

- Grostern, A., C.M. Sales, W.-Q. Zhuang, O. Erbilgin, R.E. Parales, and L. Alvarez-Cohen. 2011. Uncovering the genetic basis of 1,4-dioxane metabolism in *Pseudonocardia dioxanivorans* strain CB1190. West Coast Bacterial Physiologists Annual Meeting, Asilomar, CA (presentation)
- Grostern, A., C.M. Sales, W.-Q. Zhuang, O. Erbilgin, R.E. Parales, S. Mahendra and L. Alvarez-Cohen. 2011. Toward a genetic and biochemical understanding of bacterial 1,4-dioxane metabolism. American Society for Microbiology Annual General Meeting, New Orleans, LA. (poster)
- C.M. Sales, A. Grostern, S. Mahendra, R. Parales and L. Alvarez-Cohen. The genome sequence of *Pseudonocardia dioxanivorans* strain CB1190. American Society for Microbiology Annual General Meeting, New Orleans, LA. (poster)
- W.-Q. Zhuang, C.M. Sales, A. Grostern, X.Y. Feng, Y.J. Tang and L. Alvarez-Cohen. Elucidation of C2 compound assimilation pathways in *Pseudonocardia dioxanivorans* CB1190 using ¹³C isotopic tracer analysis. American Society for Microbiology Annual General Meeting, New Orleans, LA. (poster)
- Grostern, A., C.M. Sales, W.-Q. Zhuang, O. Erbilgin, X. Feng, Y. Tang, S. Mahendra and L. Alvarez-Cohen. 2010. A Genome-enabled investigation of 1,4-dioxane metabolism by

- the bacterium *Pseudonocardia dioxanivorans* strain CB1190. Strategic Environmental Research & Development Program Annual Symposium, Washington, DC. (poster)
- Grostern, A., C.M. Sales, S. Mahendra and L. Alvarez-Cohen. 2010. Genome assembly of the 1,4-dioxane degrading *Pseudonocardia dioxanivorans* strain CB1190. 13th International Symposium on Microbial Ecology (ISME-13), Seattle, WA. (poster)
 - Sales, C.M., A. Grostern, S. Mahendra and L. Alvarez-Cohen. 2010. Identification of monooxygenases involved in 1,4-dioxane biodegradation by *Pseudonocardia dioxanivorans* CB1190. 2010 General Meeting American Society of Microbiology, San Diego, CA. (poster)
 - Sales, C.M., S. Mahendra, A. Grostern, R.E. Parales and L. Alvarez-Cohen. 2009. Genome sequencing of the 1,4-dioxane-utilizing bacterium *Pseudonocardia dioxanivorans* CB1190. Strategic Environmental Research & Development Program Annual Symposium, Washington, DC. (poster)
 - Sales, C.M., Sharp, J.O., and L. Alvarez-Cohen. 2009. Identification of genes involved in propane-enhanced biotransformation of *N*-nitrosodimethylamine (NDMA) in two actinomycetes. 109th General Meeting of the American Society for Microbiology. Philadelphia, PA. (poster)
 - Mahendra, S., Sales, C.M., Parales, R.E., and L. Alvarez-Cohen. 2009. Genome sequencing of 1,4-dioxane utilizing bacterium *Pseudonocardia dioxanivorans* CB1190. 4th Annual DOE JGI User Meeting, Joint Genome Institute. Walnut Creek, CA. (poster)
 - Sales, C.M., Sharp, J.O., and L. Alvarez-Cohen. 2008. Identifying the enzymes responsible for biodegradation of *N*-nitrosodimethylamine in different propanotroph strains. 108th General Meeting of the American Society for Microbiology. Boston, MA. (poster)
 - Sales, C.M., Sharp, J.O., LeBlanc, J., Le J., Wood, T.K., Eltis, L.D., Mohn, W.W., and L. Alvarez-Cohen. 2007. Biodegradation of *N*-nitrosodimethylamine by bacterial propane monooxygenases. 2007 Partners in Environmental Technology Symposium & Workshop, SERDP & ESTCP. Washington, D.C. (poster)
 - Sales, C.M., Sharp, J.O., LeBlanc, J., Le J., Mohn, W.W., Eltis, L.D., Wood, T.K., and L. Alvarez-Cohen. 2007. Identifying the involvement of a propane monooxygenase in the biodegradation of *N*-nitrosodimethylamine (NDMA) in *Rhodococcus* sp. RHA1. 107th General Meeting of the American Society for Microbiology. Toronto, Canada. (poster)
 - Sales, C.M., Sharp, J.O., and L. Alvarez-Cohen. 2006. Differences in *N*-nitrosodimethylamine biodegradation kinetics between two propanotrophs. 2006 Partners in Environmental Technology Technical Symposium & Workshop, SERDP & ESTCP. Washington, D.C. (poster)

Ph.D. dissertations

- Sales, C. M. 2012. Functional genomics of bacterial degradation of the emerging contaminants 1,4-dioxane and *N*-nitrosodimethylamine (NDMA). Ph.D. thesis. University of California, Berkeley, Berkeley, CA.
- Mahendra, S. 2007. Biodegradation of 1,4-dioxane by aerobic bacteria: experimental studies and modeling of oxidation kinetics, co-contaminant effects, and biochemical pathways. Ph.D. thesis. University of California, Berkeley, Berkeley, CA.
- Sharp, J. O. 2006. Aerobic bacterial degradation of the water pollutant *N*-nitrosodimethylamine (NDMA). Ph.D. thesis. University of California, Berkeley, Berkeley, CA.